Grantham Research Institute on Climate Change and the Environment

Centre for Climate Change Economics and Policy

Believe me when I say green! Heterogeneous expectations and climate policy uncertainty

Emanuele Campiglio, Francesco Lamperti and Roberta Terranova

March 2023

Centre for Climate Change Economics and Policy Working Paper No. 419
ISSN 2515-5709 (Online)

Grantham Research Institute on Climate Change and the Environment Working Paper No. 395
ISSN 2515-5717 (Online)

Economic and Social Research Council

The Grantham Foundation for the Protection of the Environment

University of Leeds

The London School of Economics and Political Science

The Centre for Climate Change Economics and Policy (CCCEP) was established by the University of Leeds and the London School of Economics and Political Science in 2008 to advance public and private action on climate change through innovative, rigorous research. The Centre is funded by the UK Economic and Social Research Council. Its third phase started in October 2018 with seven projects:

1. Low-carbon, climate-resilient cities
2. Sustainable infrastructure finance
3. Low-carbon industrial strategies in challenging contexts
4. Integrating climate and development policies for 'climate compatible development'
5. Competitiveness in the low-carbon economy
6. Incentives for behaviour change
7. Climate information for adaptation

More information about CCCEP is available at www.cccep.ac.uk

The Grantham Research Institute on Climate Change and the Environment was established by the London School of Economics and Political Science in 2008 to bring together international expertise on economics, finance, geography, the environment, international development and political economy to create a world-leading centre for policy-relevant research and training. The Institute is funded by the Grantham Foundation for the Protection of the Environment and a number of other sources. It has 13 broad research areas:

1. Biodiversity
2. Climate change adaptation and resilience
3. Climate change governance, legislation and litigation
4. Climate, health and environment
5. Environmental behaviour
6. Environmental economic theory
7. Environmental policy evaluation
8. International climate politics
9. Science and impacts of climate change
10. Sustainable public and private finance
11. Sustainable natural resources
12. Transition to zero emissions growth
13. UK national and local climate policies

More information about the Grantham Research Institute is available at: www.lse.ac.uk/GranthamInstitute

Suggested citation:
Campiglio E, Lamperti F and Terranova R (2023) *Believe me when I say green! Heterogeneous expectations and climate policy uncertainty*. Centre for Climate Change Economics and Policy Working Paper 419/Grantham Research Institute on Climate Change and the Environment Working Paper 395. London: London School of Economics and Political Science

Believe me when I say green!
Heterogeneous expectations and climate policy uncertainty*

Emanuele Campiglio[a,b,c], Francesco Lamperti[b,d], and Roberta Terranova[b]

[a] *University of Bologna, Department of Economics, Bologna, Italy*
[b] *RFF-CMCC European Institute on Economics and the Environment (EIEE), Milan, Italy*
[c] *LSE Grantham Research Institute on Climate Change and the Environment, London, United Kingdom*
[d] *Institute of Economics and EMbeDS, Scuola Superiore Sant'Anna, Pisa, Italy*

This version: March 2023

Abstract

We develop a dynamic model where heterogeneous firms take investment decisions depending on their beliefs on future carbon prices. A policy-maker announces a forward-looking carbon price schedule but can decide to default on its plans if perceived transition risks are high. We show that weak policy commitment, especially when combined with ambitious mitigation announcements, can trap the economy into a vicious circle of credibility loss, carbon-intensive investments and increasing risk perceptions, ultimately leading to a failure of the transition. The presence of behavioural frictions and heterogeneity - both in capital investment choices and in the assessment of the policy-maker's credibility - has strong non-linear effects on the transition dynamics and the emergence of 'high-carbon traps'. We identify analytical conditions leading to a successful transition and provide a numerical application for the EU economy.

Keywords: beliefs; behavioural macroeconomics; credibility; investment decision-making; heterogeneous expectations; low-carbon transition; policy uncertainty

JEL codes: C63, D84, E71, Q54, Q58

*Email for correspondence: roberta.terranova@eiee.org. We would like to thank Louison Cahen-Fourot, Louis Daumas, Domenico Delli Gatti, Francesca Diluiso, Giorgos Galanis, George Krivorotov, Piergiuseppe Pezzoli, Simon Quemin, Alessandro Spiganti, Marko Stojanovic, Frank Venmans, as well as participants to the EAERE 2022, EAEPE 2022, ESEE 2022, GRASFI 2022, EAEPE 2022, WEHIA 2021 and 2022, SURED 2022, EPSF 2022 and OCC 2022 conferences. We would also like to thank seminar participants at the London School of Economics, University of Montpellier and Università Cattolica del Sacro Cuore. The research leading to these results has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (Grant agreement No 853050 - SMOOTH). We also acknowledge support from the Grantham Foundation for the Protection of the Environment and the UK Economic and Social Research Council through the Centre for Climate Change Economics and Policy.

1 Introduction

Transitioning to a low-carbon economy will require an expanding share of firms to allocate their physical capital investments to carbon-free technologies (IPCC, 2022). In market economies, investment choices are mainly driven by profitability expectations: firms adopt low-carbon investment strategies only if they expect them to be more convenient than carbon-intensive alternatives. In turn, a major factor affecting their relative profitability expectations is the expected strength and timing of future climate mitigation policies, the most prominent of which is the introduction of a carbon price (World Bank, 2022).

What do firms expect future carbon prices to be, and how do they formulate their expectations? It is reasonable to assume policy-makers' stated intentions to act as a key expectation anchor. For instance, a large number of countries in recent years have publicly pledged to reach 'net-zero' emissions by a certain date in the future,[1] with an implicitly associated optimal carbon price policy (Riahi et al., 2021). However, policy announcements trigger the desired behavioural changes only if individuals believe these will be followed by actual policy actions. Recent history counts many examples of failed policy commitments or complete policy reversals. In the climate/energy policy sphere, examples include the rapid and sometimes retroactive withdrawal from clean energy subsidies in Europe (Sendstad et al., 2022); the introduction and subsequent repeal of a carbon tax in Australia (Crowley, 2017); the troubled relationship of the US with the Paris Agreement (Urpelainen and Van de Graaf, 2018); the recalibration of French fiscal policy after the Gilet Jaunes movement (Douenne and Fabre, 2020); and the numerous cases of fossil fuel subsidy reform withdrawals following social unrest (McCulloch et al., 2022). Experience of policy volatility and uncertainty can lead individuals to discount the credibility of their policy-makers' commitments.

Failure to meet policy targets is often motivated by the perception of excessively high transition risks, i.e. socio-economic costs generated by the process of structural change away from carbon-intensive technologies, triggered by policy implementation, technological progress or evolving preferences (Campiglio and van der Ploeg, 2022). In the context of the low-carbon transition, certain sectors and countries are likely to lose out because of mitigation policies. More systemic effects could potentially ensue - increase in energy prices, output loss, physical capital stranding, unemployment, loss of competitiveness, financial instability - with dire political implications for the implementing policy-maker (Carattini et al., 2018; Comerford and Spiganti, 2022; Konc et al., 2022; Vona, 2018; Semieniuk et al., 2021). Governments in both high-income and emerging economies might ultimately succumb to the discontent of (parts of) the population and return on their steps, or just be replaced by new governments following different policy directions.

Confronted with multiple sources of uncertainty, individuals develop heterogeneous beliefs re-

[1] 'Net-zero' commitments have been announced by the European Union (climate neutrality by 2050), United States (2050), China (2060), Russia (2060) and India (2070), among others (Hale et al., 2022; Fankhauser et al., 2022).

garding the credibility of the policy-maker and the future schedule of carbon prices, depending on the information they have access to, their ability to process it, their political preferences, the length of their planning horizon, and a number of other behavioural factors. Beliefs and expectations will also change in time with the arrival of new information, possibly amplified by herd behaviour dynamics or, to the contrary, restrained by deep-rooted convictions creating inertia in the updating process. Indeed, in the scarce empirical evidence we have on the matter, expectations about both future climate risks and climate policies - crucial in shaping investments towards alternative technological options - have been shown to be volatile and heterogeneous across economic agents (Barradale, 2014; Engle et al., 2020; Krueger et al., 2020; Noailly et al., 2021; Nordeng et al., 2021; Stroebel and Wurgler, 2021).

In this paper, we develop a dynamic model to study how heterogeneous and dynamic beliefs can affect the low-carbon transition, and how climate mitigation policies should be appropriately communicated and implemented.[2] In our model, firms choose the proportion of investments to allocate between low- and high-carbon technological options depending on their expected relative profitability, itself a function of expectations of future carbon prices. To formulate their carbon price expectations, firms observe the forward-looking carbon price schedule announced by the policy-maker. We assume firms to belong to one of two populations: 'believers' trust the climate mitigation commitment of the policy-maker more than 'sceptics'. Policy-makers can deviate from their carbon price plans if they are concerned of transition socio-economic costs, which they perceive to be larger when the economic system is more carbon-intensive and when the carbon price is higher. In each period, firms can decide to switch their policy beliefs on the basis of the government's track-record in keeping its word. The share of firms believing the policy-maker's announcements can thus be interpreted as a measure of the policy-maker's credibility, which is endogenous to past policy choices. We allow firms' choices - how to allocate investments across technologies and whether to believe in the policy-maker announcements - to be subject to behavioural frictions. By behavioural frictions, we mean factors such as bounded rationality, cognitive limitations, incomplete information and any other behavioural dimension preventing firms from immediately choosing the optimal alternative among their available options. As a result, these multiple frictions lead to heterogeneous choices across firms. More specifically, building on the literature on behavioural macroeconomics (see Hommes, 2021), we introduce parameters for both the *investment* and *belief responsiveness* of firms, i.e. their ability to rapidly incorporate new available information and, if appropriate, switch to different investment/belief strategies. We explore the full range of possible values for these behavioural parameters, going from zero (investment choices are made at random, as strong frictions fully debilitate firms' decision-making process) to infinity (the 'neoclassical limit': all firms immediately make the marginally most convenient choice, with no frictions).

[2]The code to replicate the results of our model is available at https://github.com/SMOOTH-ERC/believe_me_green.

We first derive analytical conclusions from a reduced version of the model. We show that, in the neoclassical limit without behavioural frictions and heterogeneity, two steady states (each fully dominated by one of the two technologies) can exist depending on i) the announced policy stringency; and ii) the policy-maker commitment level. We derive the conditions for existence of these steady states and find that the combination of ambitious mitigation plans and a weakly-committed policy-maker can lead to the emergence of multiple equilibria (i.e. a 'high-carbon trap'). When we introduce behavioural frictions and the associated heterogeneity of beliefs/expectations, we identify a set of 'behavioural premiums' that modify the long-run equilibria of the system. The conditions for existence of the low-carbon steady state become harder to satisfy, with a higher minimum tax target compared to the neoclassical limit case, and the emergence of a new minimum commitment requirement. However, we also find that behavioural frictions could help the unambitious policy-maker to achieve 'mid-carbon' steady states. Finally, we identify the sufficient conditions, given commitment and belief responsiveness levels, for the tax announcement to create a unique low-carbon steady state.

We then calibrate the full version of the model to European data and run forward-looking numerical simulations. We distinguish two scenarios: (i) full commitment by the policy-maker, with climate policy targets always met regardless of transition costs; and (ii) less-than-full climate policy commitment. Under a fully-committed policy-maker and a sufficiently strong tax announcement, the economic system is almost always eventually reaching full decarbonisation, but the speed of the transition is significantly affected by behavioural dimensions (i.e. the investment and belief responsiveness of the firms' population). When allowing for the policy-maker to default on its commitment due to potential transition costs, we find that the decarbonisation can endogenously fail, getting trapped into a vicious circle of credibility loss, carbon-intensive investment choices and increasing transition risk perceptions. Our results suggest that, while the weakly-committed policy-maker could succeed in purposely overshooting its policy targets so to push the transition through before the credibility loss takes over, exceeding in deception can backfire and eventually compromise the transition process. Finally, we explore the role of belief polarisation, i.e. the distance between the belief systems of believers and sceptics, finding that, under certain conditions, a highly polarised belief system can lead to a more rapid transition than one with mildly polarised beliefs.

We build upon and contribute to three broad streams of research, which have largely proceeded independently so far. First, we build on the literature investigating the effects of climate policy uncertainty and the importance of 'credible' climate policies in achieving a rapid and orderly transition to a carbon-free economy (e.g. Barradale, 2014; Battiston et al., 2021; Berestycki et al., 2022; Bosetti and Victor, 2011; Dunz et al., 2021; Fried et al., 2022; Fuss et al., 2008; Helm et al., 2003; Nemet et al., 2017; van der Ploeg and Rezai, 2020). Against such background, we introduce the innovation of heterogeneous forward-looking agents dynamically updating their beliefs about future climate policy in response to the behaviour of the government. Second, our modelling framework, where policy expectations affect firms' investment decisions and these in turn affect actual policy implementation, connects to the literature on policy time inconsistency (Kydland and Prescott, 1977; Barro and Gordon, 1983). Economic modelling in this area has usually been done in a context of homogeneous and rational expectations, with the aim of identifying optimal policy paths. We instead adopt an heterogeneous beliefs framework wherein agents continuously adapt their choices. Third, we contribute to the modelling literature investigating the role of bounded rationality, such as the finiteness of forward-looking planning horizons (Spiro, 2014; Quemin and Trotignon, 2021) or the formation of heterogeneous and systematically biased expectations (Bordalo et al., 2022; Evans and Honkapohja, 2012; Gigerenzer and Brighton, 2009; Hommes, 2006). In particular, we draw upon the modelling framework with heterogeneous and dynamic expectations developed by Brock and Hommes (1997, 1998), itself rooted in discrete choice theory (McFadden, 1974).

This family of logit models has been fruitfully applied to a number of research questions in monetary economics, especially those related to the interaction between heterogeneous inflation expectations and monetary policy decisions (Evans and Honkapohja, 2006; Salle et al., 2013; De Grauwe and Macchiarelli, 2015; Hommes and Lustenhouwer, 2019; Assenza et al., 2021).[3] The framework is well suited to capture decisions taken by boundedly rational agents who strive to choose the best option but are not always able to do so, for a number of unobservable factors. Its applications to climate-related questions has been limited so far. Annicchiarico et al. (2022) introduce the possibility for agents to switch among forecasting rules for output and inflation and study how this affects the impact of climate policies in a New Keynesian model. Zeppini (2015) and Mercure (2015) apply a similar logit framework to the choice on technology adoption, while Cafferata et al. (2021) and Dávila-Fernández and Sordi (2020) focus on switching attitudes towards green policies, and Galanis et al. (2022b) employs it to study country participation in international environmental agreements. Cahen-Fourot et al. (2022) study how heterogeneous transition expectations affect investment decision choices in a forward-looking probit model with capital 'stranding'. Torren-Peraire et al. (2023) study the interaction between decarbonisation and cultural change in an agent-based model where the influence individuals have on each other recalls Brock and Hommes (1998) discrete choice models. We advance this nascent literature by developing a double logit framework - including both a backward-looking choice on the perceived ex-post policy-maker's credibility and a forward-looking choice on the technology to invest in - to study the dynamic interactions between climate policy strategies, firms' behavioural response to such policy choices and the transition to a low-carbon economy.

The remainder of the paper is structured as follows. Section 2 presents the model. Section 3 derives some analytical conclusions using a reduced version of the model. Section 4 explains our calibration strategy for the full model. Section 5 presents and discusses our numerical results under the assumption of a fully committed policy-maker. Section 6 extends the numerical analysis to the

[3] Probit models can also be fruitfully employed in this context, see Galanis et al. (2022a).

case of a weakly committed policy-maker. Section 7 concludes.

2 The model

We consider an economy in discrete time, moving from t_0 to T. The system is populated by a continuum of firms producing a homogeneous final good Y. We assume demand of good Y to grow at an exogenous rate g_Y and supply to always be able to satisfy demand thanks to expansions of the capital stock, which is the only factor of production.[4] Two types of capital stocks exist: (i) high-carbon (fossil-fuelled) capital K_h, producing greenhouse gas emissions; and (ii) low-carbon capital K_l, with no production of emissions. Aggregate capital stock is the sum of the two technology-specific stocks, i.e. $K = \sum K_i$, where $i \in \{l, h\}$ denotes the two technologies. We define κ as the share of low-carbon capital over the aggregate capital stock, i.e. $\kappa_t \equiv \frac{K_{l,t}}{\sum_i K_{i,t}}$.

The following sequence of events takes place in the model: (i) at t_0, the policy-maker announces a future schedule of target carbon prices; (ii) at each time $t \geq t_0$, the policy-maker implements a carbon price, which may or may not be the one previously announced; (iii) at each time $t \geq t_0$, firms evaluate the credibility of the policy announcement and take investment decisions.

2.1 Policy announcements and expectation dynamics

Climate policy in our model takes the form of a tax on high-carbon capital production. At t_0, the policy-maker announces its intention to implement a schedule of rising carbon tax rates $\bar{\tau}_t$, starting from an exogenous level τ_0 and increasing at a constant growth rate $\bar{g}_\tau$.[5] That is, at t_0 the policy-maker announces that, at each $t \in [t_0, T]$,

$$\bar{\tau}_t = \tau_0(1 + \bar{g}_\tau)^t. \tag{1}$$

At each time t, firms formulate expectations on future carbon prices by 'discounting' the policy-maker announcement. We assume firms to be either 'believers' (b) or 'sceptics' (s), with $j \in \{b, s\}$ indicating the belief type. Belief-specific carbon price expectations are defined as

$$E_{j,t}(\tau_{t+r}) = \tau_0(1 + \epsilon_j \bar{g}_\tau)^{t+r}, \tag{2}$$

where the operator $E_t(\cdot)$ denotes the expectations formulated at time t and $\epsilon_j \in [0, 1]$ is a parameter indicating to what extent firms believe to the policy announcement, with $\epsilon_b > \epsilon_s$ (i.e believers trust

[4] Alternatively, without loss of generality, one might consider a linear production function where labour and capital are complements, and the former is abundant. We abstract from technical change.

[5] There is ample consensus in the climate economics literature that carbon prices should increase to stimulate emission mitigation and that, in many cases, the optimal price trajectory is exponential (e.g. Golosov et al., 2014; Nordhaus, 2017; Riahi et al., 2021; van den Bremer and van der Ploeg, 2021). For an alternative approach leading to declining optimal carbon prices, see Daniel et al. (2019).

policy announcements more than sceptics). The distance between ϵ_s and ϵ_b can be interpreted as a proxy for opinion polarisation: when $\epsilon_s = \epsilon_b$, sceptics' expectations are entirely aligned to those of believers (homogeneous population); when $\epsilon_s = 0$ and $\epsilon_b = 1$, the system of beliefs is instead heavily polarised.[6].

Similarly to Brock and Hommes (1997, 1998), firms switch belief type depending on their relative accuracy in predicting the policy-maker's behaviour in the past. The larger is the difference between forecast and actual policy, the higher is the likelihood of firms changing their belief. We define the fitness measure of expectation rule j in period t, $U_{j,t}$, as the weighted sum of the last observed absolute prediction error, and the previous value of the fitness measure:

$$U_{j,t} = (1-\eta)\left|E_{j,t-1}(\tau_t) - \tau_t\right| + \eta U_{j,t-1}, \tag{3}$$

where $0 \leq \eta \leq 1$ is a memory parameter indicating to what extent firms update their evaluation of the expectation rule with new information. If η is set to zero, firms only consider the last prediction error, whereas if $\eta > 0$, firms also take into account past values of the fitness measure.

The switching mechanism between beliefs, based on the accuracy of predictions evaluated in the previous period, determines the share $n \in [0,1]$ of firms adopting belief b (i.e. choosing to be 'believers') at time t:

$$n_t = \frac{\exp(-\beta U_{b,t-1})}{\sum_j \exp(-\beta U_{j,t-1})}, \tag{4}$$

where $\beta \geq 0$ represents 'belief responsiveness', i.e. the responsiveness of firms' beliefs to prediction errors and, consequently, to policy choices.[7]. Low values of β indicate that firms' beliefs react mildly to prediction errors, with $\beta = 0$ indicating a population of firms evenly split between the two beliefs, regardless of U_j. If instead β tends to infinity, all firms immediately adopt the best performing expectation rule, even if the performance gap between the two is small (i.e. a 'bang-bang' solution) . In our framework, since firms' expectations are driven by announced and implemented policies, β also indicates the speed of firms' response to policy choices. We interpret weak belief responsiveness as a consequence of high behavioural frictions, which prevent firms to promptly react to new information.[8] Based on past policy choices and firms' response to them, the share of believers n_t is determined endogenously and measures the ex-post policy-maker's credibility, similarly to Hommes and Lustenhouwer (2019).

[6]When using the expression 'opinion polarisation', we refer to the difference between the two system of beliefs, regardless of the size of the two populations.

[7]In the framework developed by Brock and Hommes (1997, 1998), β is referred to as the 'intensity of choice' parameter.

[8]Appendix A shows that, under certain assumptions, the rule illustrated in equation (4) reflects a continuous distribution of tax expectations. According to this interpretation, β represents the inverse of the dispersion of such distribution.

2.2 Investment choices and capital dynamics

At each time t, the aggregate amount of investments can be expressed as a function of the existing capital stock, given exogenous and constant growth rate of output g_Y (also equivalent to growth rate of the capital stock) and capital depreciation rate δ:

$$I_t = (g_Y + \delta)K_t. \tag{5}$$

Building on their chosen beliefs, firms decide how to allocate their investments I across the two available technologies. They do so by evaluating the net present value of the expected production costs $\Theta_{i,t}$ associated to each technology $i \in \{l, h\}$ as

$$E_{j,t}(\Theta_{i,t}) = \begin{cases} \sum_{r=1}^{R} D^r \theta_{i,t+r} & \text{if } i = l, \\ \sum_{r=1}^{R} D^r \theta_{i,t+r} \left[1 + E_{j,t-1}(\tau_{t+r})\right] & \text{if } i = h, \end{cases} \tag{6}$$

where R is the length of firms' planning horizon, $D = \frac{1}{1+\rho}$ is a discount factor with discount rate ρ, and θ is the technology-specific pre-tax production cost, comprising both capital installation and operational costs. Once a firm, pertaining to a specific belief population j, has assessed the cost prospects of the two technologies, it allocates its investments to the technology it perceives to be the most convenient.

Not all firms pertaining to the same belief population will invest in the same technology. In line with the discrete choice theoretical framework, we assume that a number of other unobservable variables affect firms' investment decisions. In addition, similarly to the belief adoption rule described in section 2.1, firms might be subject to imperfect information, bounded rationality and a number of behavioural frictions.

We can thus define the j-specific share of low-carbon investment $\chi_{j,t} \equiv \frac{I_{l,j,t}}{\sum_i I_{i,j,t}}$, as

$$\chi_{j,t} = \frac{\exp(-\gamma E_{j,t}(\Theta_{l,t}))}{\sum_i \exp(-\gamma E_{j,t}(\Theta_{i,t}))} \tag{7}$$

where γ indicates 'investment responsiveness', that is, the responsiveness of firms' investment to the difference in expected costs. As cost expectations depend on policy announcements and implementation, investment responsiveness, similarly to belief responsiveness β, can be interpreted as the speed of firms' response to the policy-maker's behaviour. Also, it signals the degree of behavioural frictions affecting firms' decisions: $\gamma = 0$ implies that firms randomly choose the technology, leading to a clean investment share of 0.5, regardless of cost differentials; if instead γ tends to infinity, all firms choose the technology they expect to be marginally more convenient. This choice is belief-specific; that is, given their different carbon price expectations, believers ad sceptics might have a different perception of which technology is the best performing.

Building on firms' technology choice (equation 7) and on the belief switching dynamics (equation 4), we can derive the evolution of the aggregate low-carbon investment share over time, χ_t, as a weighted average of $\chi_{j,t}$

$$\chi_t = n_t\chi_{b,t} + (1 - n_t)\chi_{s,t}. \tag{8}$$

Finally, assuming the standard capital law of motion $K_{i,t+1} = (1 - \delta)K_{i,t} + I_{i,t}$, and building on equation (5), we can define the evolution of the low-carbon share of capital κ as a function of χ:

$$\kappa_t = \frac{\kappa_{t-1}(1 - \delta) + \chi_t(g_y + \delta)}{1 + g_y}. \tag{9}$$

2.3 Transition risks and policy implementation

Once investment decisions are taken, the policy-maker decides if and to what extent the announced policy will actually be implemented. While the policy-maker's intentions might have been sincere at the time of the announcement, the potential costs related to the low-carbon transition, together with their electoral implications, might weaken its resolution and lead to a change of its mitigation strategy. Transition costs can be of various nature, ranging from higher energy bills for households and firms to unemployment in high-carbon industries and systemic financial disruptions.

We assume our policy-maker to formulate a 'transition risk index' $0 \leq \pi \leq 1$, increasing in both the announced tax target $\bar{\tau}$ and in the carbon intensity of the economy's productive basis, $1-\kappa$. The intuition is that governments will consider their economies more exposed to the risk of transition costs when climate objectives are more stringent (i.e. higher announced carbon prices) and when a larger part of the productive system relies on high-carbon technologies (e.g. oil and gas extraction; coal- and gas-fuelled electricity production; internal combustion engine vehicle production; etc.). This is in line with the approach used by Peszko et al. (2020) to calculate the index of countries' exposure to low-carbon transition risks.[9] When no policy is scheduled ($\bar{\tau} = 0$), or if the economy is already entirely decarbonised ($\kappa = 1$), we assume perceived transition risks to be equal to zero and to have no weight in the policy-maker's decisions. We thus write

$$\pi_t = 1 - \frac{1}{1 + a(1 - \kappa_t)\bar{\tau}_t}, \tag{10}$$

where a is a parameter capturing the vulnerability of the economy (as perceived by the policy-maker) to transition risks. This vulnerability might depend on several factors, such as the exposure of the banking and financial system to transition risks, the fragility of the welfare system and the vulnerability to social turmoil (see Appendix D for a graphical representation). One can interpret a as the inverse of the index of resilience to low-carbon transition impacts proposed by Peszko et al.

[9] More in detail, Peszko et al. (2020) uses four measures to compute the transition risk exposure index: carbon intensity of manufacturing exports; committed power emissions as a proportion of current annual power generation; fossil fuel export as a proportion of GDP; expected resource rents as a proportion of GDP.

(2020). While our concave functional form is the one we believe to be most representative of the current debate on transition costs, where even the announcement of relatively mild mitigation policies can lead to large protests by vocal minorities, we test for alternative formulations in Appendix D, showing that the qualitative results of the model remain untouched.

Once transition risks have been estimated, the policy-maker weighs them against its climate mitigation objectives in order to choose the carbon tax to actually implement. We formalise this policy choice as a weighted average between the announced tax target $\bar{\tau}_t$ and the tax target reduced by the transition risk index, $\bar{\tau}_t(1-\pi_t)$:

$$\tau_t = c\bar{\tau}_t + (1-c)\bar{\tau}_t(1-\pi_t), \tag{11}$$

where $0 \le c \le 1$ is an exogenous parameter indicating the policy-maker commitment to climate mitigation objectives. A value of $c = 1$ indicates a policy-maker fully committed to mitigation, who will therefore always impose taxes as scheduled, regardless of the transition risks involved. On the other hand, $c = 0$ represents the case of a policy-maker fully committed to the reduction of transition risks and willing to reduce the tax to a floor level of $\bar{\tau}_t(1-\pi_t)$.[10]

3 Analytical results

In this section, we explore the analytical properties of the model. We study the existence of multiple steady states and characterise their stability in order to assess the long-term behaviour of the low-carbon transition and its dependence on policy settings and firms' attitudes. We start by introducing three sets of assumptions that grant us tractability without affecting the qualitative behaviour of the system.[11]

Assumption 1. *The announced carbon price is positive and constant, i.e.* $\bar{\tau}_t = \bar{\tau} \in R_0^+$ *for all* t.

We assume the policy-maker to just announce a carbon tax $\bar{\tau}$, without including a forward-looking increasing schedule of prices. In other words, a discrete jump from τ_0 (the initial tax level) and $\bar{\tau}$ is announced. The actual tax τ_t might however be different from what announced depending on transition risk perceptions and the policy-maker's commitment level, as in the full model. This setting simplifies our analysis by making our dynamic model autonomous, i.e. independent of time.

[10] The model outlined here can also be formulated using Bayesian updating. In this case, one can think of firms aiming to infer the policy-maker's commitment level by observing the implemented tax, which is subject to noise, as described by: $\tau_t = c\bar{\tau} + (1-c)\bar{\tau}(1-\pi_t) + \epsilon_t^\tau$. Initially, firms hold heterogeneous priors, denoted by $F_0 = N(\hat{c}_0, \sigma_{c_0})$, where $\hat{c}_0$ represents the prior mean estimate of c and σ_{c_0} the spread of the prior. These priors are then updated based on the following process: $F_t = N\left(\frac{1}{\frac{1}{\sigma_{c_0}^2}+\sum_{s=1}^t \frac{1}{\sigma_{c_s}^2}}\left(\frac{\hat{c}_0}{\sigma_{c_0}^2} + \sum_{s=1}^t \frac{\hat{c}_s}{\sigma_{c_s}^2}\right), \left(\frac{1}{\frac{1}{\sigma_{c_0}^2}+\sum_{s=1}^t \frac{1}{\sigma_{c_s}^2}}\right)^{-1}\right)$, where $\hat{c}_s = \frac{\tau_s - \bar{\tau}_s(1-\pi_s)}{\tau_s \pi_s}$ and $\sigma_{c_s} = \frac{\sigma_\epsilon^\tau}{\tau_s \pi_s}$.

[11] In sections 5 and 6 we will remove these assumptions and run numerical simulations for the full calibrated model.

At the same time, however, it allows us to obtain more general results compared to the case of a tax target growing over time according to a specific function.

Assumption 2. *We set $\delta = 1$; $\eta = 0$; $\epsilon_b = 1$; and $\epsilon_s = 0$.*

To simplify the derivation of analytical results, we assume full capital depreciation, i.e. $\delta = 1$. While this assumption affects the dynamics of the model, it does not modify steady states, as these are independent of the rate of capital depreciation. We set the memory parameter $\eta = 0$, thereby assuming agents to fully update the fitness measure of beliefs with the newly available information. Finally, we set $\epsilon_s = 0$ and $\epsilon_b = 1$, so to characterise the two beliefs types to their extreme version (i.e. believers entirely believe in policy announcements; sceptics don't believe in them at all). i.e. sceptics will expect the tax rate τ never to move from its initial level τ_0, while believers will expect it to be equal to the announced rate $\bar{\tau}$.

Assumption 3. *We impose $\tau_0 < \frac{\theta_l - \theta_h}{\theta_h}$; $\bar{\tau} > \frac{\tau_0}{1 - a\tau_0}$; and $\tau_0 < \frac{1}{a}$.*

The first condition on τ_0 implies that the starting tax rate does not cover for the percentage cost difference between low- and high-carbon technologies. The second and third conditions assume sceptics' prediction errors to be positive, i.e. sceptics tend to underestimate actual climate policy, independently of c and k.[12]

Under these assumptions, the dynamics of the low-carbon capital share is given by

$$\kappa_{t+1} = (\chi_b - \chi_s)n_{t+1} + \chi_s \equiv f(\kappa_t), \tag{12}$$

where χ_b and χ_s are independent of time because the tax target is now a constant (see Assumption 1). The share of believers n_{t+1} depends on the carbon intensity of capital stock κ_t as follows:

$$n_{t+1} = \left[1 + \exp\left(-\beta\left\{2\bar{\tau}\left[c + \frac{1-c}{1 + a(1-\kappa_t)\bar{\tau}}\right] - \tau_0 - \bar{\tau}\right\}\right)\right]^{-1}. \tag{13}$$

Since $n \in [0, 1]$ and $\chi_b > \chi_s$, $f(\kappa)$ is bounded between the low-carbon investment shares of the two populations, χ_s and χ_b, both of which are bounded between 0 and 1. Combining equations 12 and 13, we can derive the following proposition:

Proposition 1. *The system has either one stable steady state or an odd number of steady states; in the latter case those with an odd index are stable and the others are unstable.*

Proof. Proof of proposition 1 is provided in Appendix B.1.

We start by evaluating $f(\kappa)$ in three illustrative cases wherein the model shows a unique stable steady state. First, we assume a fully committed policy-maker ($c = 1$), implying that the tax

[12] Note that, by construction, believers' prediction errors are either zero or negative.

actually implemented equals the tax target in every t, independently of κ. In this case, $f(\kappa) = (\chi_b - \chi_s)\left[1 + \exp\left(-\beta\left(\bar{\tau} - \tau_0\right)\right)\right]^{-1} + \chi_s$. Second, we set $\beta = 0$, that is, firms choose their belief type at random. This leads to $n = 0.5$ and thus $\kappa = \frac{1}{2}(\chi_b + \chi_s)$ at every point of time. Third, we assume $\gamma = 0$, i.e. believers and sceptics split their investment equally between high- and low-carbon technologies, i.e. $\chi_b = \chi_s = 0.5$ at each point of time. The resulting equilibrium level of the low-carbon capital share is $f(\kappa) = 0.5$.

In the more general case where the policy-maker is not fully committed ($c < 1$) and both belief and investment responsiveness are not equal to zero ($\beta > 0$ and $\gamma > 0$), the system may present multiple steady states. To explore them, we consider two settings that exhaustively describe agents' behavioural attitudes: (i) a benchmark case, where belief and investment responsiveness are infinite, i.e. $\beta = \gamma = +\infty$; and (ii) a scenario characterised by boundedly rational decisions, with finite β and γ. The first scenario proxies what we shall call a *neoclassical limit* case – mirroring the label used in Brock and Hommes (1997) – wherein all agents choose the superior belief type and the cheapest technology at each time step, even if by a small margin. In contrast, the second scenario includes *behavioural frictions* inversely proportional to β and γ (see more details in Appendix A).

3.1 Steady states in the neoclassical limit

In the neoclassical limit scenario ($\beta = \gamma = \infty$) the policy-maker can induce, via its announcements and level of commitment, three different types of systems, characterised by: i) a unique high-carbon steady-state; ii) a unique low-carbon steady state; or iii) multiple steady states with varying basins of attraction. The two lines depicted in Figure 1 illustrate the thresholds on the tax target (black line) and on the commitment level (blue line) delimiting the parameter spaces in which low- and high-carbon steady states exist. Such conditions of existence are outlined in the following proposition.

Proposition 2. *Under the assumption that $\beta = \gamma = \infty$,*

(i) The low-carbon steady state $\kappa_l^* = 1$ exists if

$$\bar{\tau} > \frac{\theta_l - \theta_h}{\theta_h} \equiv \bar{\tau}^*; \tag{14}$$

(ii) The high-carbon steady state $\kappa_h^* = 0$ exists if

$$\bar{\tau} < \bar{\tau}^* \qquad \text{or} \qquad c < \frac{1}{2} - \mu_1 \equiv c^* \tag{15}$$

where $\mu_1 = \frac{\bar{\tau} - \tau_0(1 + a\bar{\tau})}{2a\bar{\tau}^2} > 0$.

Proof. Proof of proposition 2 is provided in Appendix B.2.

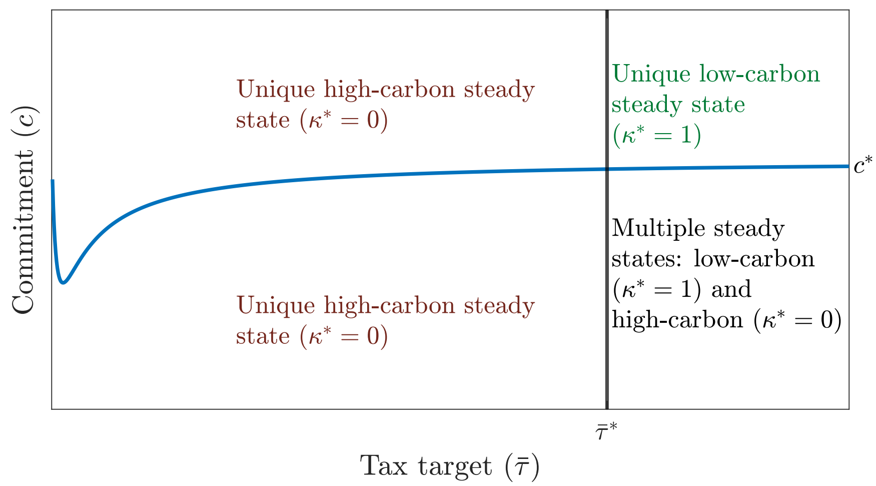


Figure 1: Existence of low- and high-carbon steady states depending on commitment c and tax target $\bar{\tau}$ in the neoclassical limit case.

Condition (14) states that the low-carbon steady state exists if the announced tax target is higher than the percentage cost difference between low- and high-carbon production costs ($\bar{\tau}^*$, i.e. the black line in Figure 1). The high-carbon steady state, instead, exists if either the announced tax target does not cover for the above mentioned cost difference $\bar{\tau}^*$, or if the policy-maker's commitment is sufficiently low (condition (15)). More specifically, the commitment threshold c^* (blue line in Figure 1) is composed of two terms. The first one, 1/2, is the upper bound of c^* and corresponds to a commitment equally split between meeting policy targets and reducing transition risks. The second term (μ_1) is decreasing in both the initial tax rate τ_0 and in the vulnerability to transition risks a. An increase in these two variables thus moves up the commitment threshold, expanding the parameter space where the high-carbon steady state exists. As evident in Figure 1, the effect of the announced tax target $\bar{\tau}$ is instead non-linear, decreasing up to a value $\bar{\tau} = \frac{2\tau_0}{1-a\tau_0}$ and increasing afterwards.[13] Since the turning point takes place for very low values of $\bar{\tau}$, we can state that, for sufficiently high and reasonable values of announced carbon prices, an increase in ambition increases the likelihood of a high-carbon steady state emerging.

By combining the conditions outlined in Proposition 2, we obtain the different long-term states of the model shown in Figure 1. A full decarbonisation of the economy (top-right quadrant) can be achieved by announcing a tax able to compensate for the relative backwardness of low-carbon technologies and by being sufficiently committed to such target. This happens because, with infinite investment responsiveness, even the smallest expected cost difference in favour of low-carbon capital

[13] To have c^* increasing in $\bar{\tau}$, $\tau_0 < \frac{1}{a}$ also needs to be verified, which is always the case in our numerical examples.

leads believers to fully invest in the clean technology ($\chi_b = 1$). Moreover, under infinite belief responsiveness and a sufficiently high commitment, the policy-maker is able to convince every firm of its credibility ($n = 1$). Together, investment and belief dynamics determine the uniqueness of the low-carbon steady state.

To the contrary, an excessively low carbon tax target ($\tau < \tau^*$) generates a unique and stable high-carbon steady state (top- and bottom-left quadrants). Intuitively, in absence of behavioural frictions, if the carbon tax is insufficient to cover the cost advantage of the high-carbon technology, all firms will avoid investing in the inferior technology, leaving $\kappa^* = 0$ to be the only possible equilibrium level of low-carbon capital.

Finally, when the announced tax target is sufficiently high but the policy-maker is weakly committed to the decarbonisation process, the system exhibits multiple equilibria (bottom-right quadrant). While announcing ambitious policy targets stimulates believers' clean investment share χ_b, not meeting such targets hampers the policy-maker's credibility (decrease in n) and therefore the decarbonisation process. This results in the emergence of multiple equilibria, generating what we label as a *high-carbon credibility trap* for the economic system.

3.2 Steady states under behavioural frictions

We now move to the analysis of the scenario characterised by behavioural frictions and hence by heterogeneity in investment decisions. In such a context, the system never achieves a full decarbonisation ($\kappa = 1$) as, even under the most favourable conditions for low-carbon technologies, a small but positive amount of capital investments will flow to high-carbon capital. The low-carbon steady state under behavioural frictions can thus be defined as $\kappa_l^* = 1 - \lambda_l$, where λ_l represents the distance between the low-carbon steady state and full decarbonisation. Similarly, the high-carbon steady state will not be equal to zero, as a minor proportion of firms will always invest in low-carbon capital or believe in the policy-maker. We hence define $\kappa_h^* = \chi_s + \lambda_h$, where λ_h indicates the distance between the high-carbon steady state and sceptics' clean investment share. Similarly to the neoclassical limit case, we prove the following proposition illustrating the conditions of existence of the low- and high-carbon steady states.

Proposition 3. *Under the assumption that β and γ are finite,*

(i) A low-carbon steady state $\kappa_l^* = 1 - \lambda_l$ exists if a positive real number $\tilde{\lambda}_l$ exists such that

$$\bar{\tau} > \frac{\theta_l - \theta_h}{\theta_h} + \nu_{\tau l} \equiv \bar{\tau}^{**} \qquad \textit{and} \qquad c > \frac{1}{2} - \mu_2 + \nu_{cl} \equiv c^{**} \tag{16}$$

where

$\tilde{\lambda}_l = \lambda_l + \varepsilon_l$, *with ε_l a small positive number and $\tilde{\lambda}_l \in (0, \frac{1}{2})$,*

$$\nu_{\tau l} = -\ln\left(\frac{\tilde{\lambda}_l}{1-\tilde{\lambda}_l}\right)\rho\{\gamma\theta_l(1+\rho)\left[1-(1+\rho)^{-(R+1)}\right]\}^{-1}$$
$$\nu_{cl} = -\ln\left(\frac{\chi_b-1+\tilde{\lambda}_l}{1-\tilde{\lambda}_l-\chi_s}\right)(2\bar{\tau}\beta)^{-1}\left(1+\frac{1}{a\tilde{\lambda}_l\bar{\tau}}\right), \textit{ and}$$
$$\mu_2 = \frac{\bar{\tau}-\tau_0(1+a\tilde{\lambda}_l\bar{\tau})}{2a\tilde{\lambda}_l\bar{\tau}^2} > 0.$$

(ii) A high-carbon steady state $\kappa_h^* = \chi_s + \lambda_h$ *exists if a positive real number* $\tilde{\lambda}_h$ *exists such that*

$$c < \frac{1}{2} - \mu_3 + \nu_{ch} \equiv c^{***} \tag{17}$$

where

$$\tilde{\lambda}_h = \lambda_h + \varepsilon_h, \textit{ with } \varepsilon_h \textit{ a small positive number and } \tilde{\lambda}_h \in (0, \chi_b - \chi_s),$$
$$\nu_{ch} = -\ln\left(\frac{\chi_b-\chi_s-\tilde{\lambda}_h}{\tilde{\lambda}_h}\right)(2\bar{\tau}\beta)^{-1}\left\{1+\frac{1}{a[1-(\chi_s+\tilde{\lambda}_h)]\bar{\tau}}\right\}, \textit{ and}$$
$$\mu_3 = \frac{\bar{\tau}-\tau_0\{1+a[1-(\chi_s+\tilde{\lambda}_h)]\bar{\tau})\}}{2a[1-(\chi_s+\tilde{\lambda}_h)]\bar{\tau}^2} > 0$$

Proof. Proof of proposition 3 is provided in Appendix B.3.

Differently from the neoclassical limit scenario, the existence of the low-carbon steady state is now subject to two simultaneous conditions (16). The first condition is that the policy target needs to be sufficiently large (i.e. $\bar{\tau} > \bar{\tau}^{**}$, or to the right of the black line in Figure 2). This condition, which guarantees that believers invest mostly in the low-carbon asset, corresponds to its analogue in the neoclassical limit case (14), but the tax threshold under behavioural frictions (τ^{**}) now has an additional term, $\nu_{\tau l}$. The presence of this 'behavioural premium' means that a tax target covering only the technology cost differential would now be insufficient to convince investors to decarbonise. The term $\nu_{\tau l}$ is positive and decreasing in γ, ρ, R, θ_h and λ_l. In other words, the threshold that the announced carbon price $\bar{\tau}$ needs to satisfy to allow for the existence of a low-carbon steady state becomes lower if the investment responsiveness γ increases, moving closer to the neoclassical limit without behavioural frictions. The threshold becomes instead harder to satisfy, the more myopic agents are in their investment choices (short planning horizon R and high discount rate ρ); the closer is the desired low-carbon steady state to full decarbonisation (low λ_l); and, similarly to the neoclassical limit case, the lower are high-carbon technology costs (low θ_h).[14]

The second condition for the existence of the low-carbon steady state in (16) is that the policy-maker's commitment needs to be sufficiently high (i.e. $c > c^{**}$, or above the green line in Figure 2); this condition ensures that most firms believe in policy targets. Similarly to c^* in (15), the threshold c^{**} has an upper bound at $\frac{1}{2}$. The second term, μ_2, is equivalent to the term μ_1 in (15), although

[14] The constraint that $\tilde{\lambda}_l < 0.5$ implies that the desired long-term capital structure is one where low-carbon capital is more abundant than high-carbon capital (i.e. $\kappa* > 0.5$). For values of $\tilde{\lambda}_l > 0.5$, where the equilibrium clean capital share can be below 50%, there is an additional constraint to be satisfied ($\tau_0 < (\theta_l - \theta_h)/\theta_h - \nu_{\tau l}$). Moreover, under $\tilde{\lambda}_l > 0.5$, the effect of γ, ρ, R, and θ_h, on the tax target threshold changes sign. The effect of $\tilde{\lambda}_l$ on the tax target threshold, instead, remains negative.

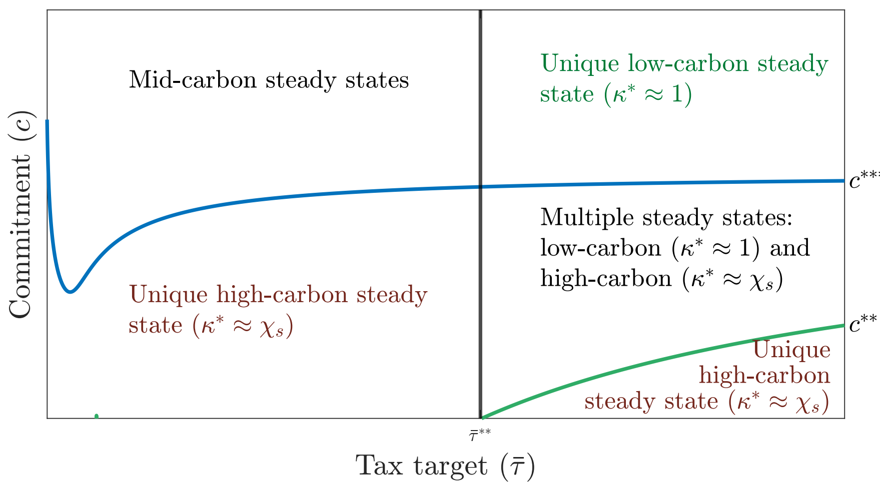


Figure 2: Existence of low- and high-carbon steady states depending on commitment c and tax target $\bar{\tau}$ in the behavioural frictions case.

augmented by a term $\tilde{\lambda}_l$, while the third term ν_{cl} can be interpreted as an additional 'behavioural premium' on commitment levels. For sufficiently small values of $\tilde{\lambda}_l$ ($\tilde{\lambda}_l < 1 - \frac{\chi_b+\chi_s}{2}$), higher behavioural frictions (i.e. a lower β) increase the necessary commitment threshold, diminishing the area of existence of the low-carbon steady state. Likewise, lower clean investment shares for both sceptics (χ_s) and believers (χ_b) increase the commitment threshold c^{**}. The overall effect of the other parameters is instead harder to establish as they might have contrasting effects on the commitment thresholds.

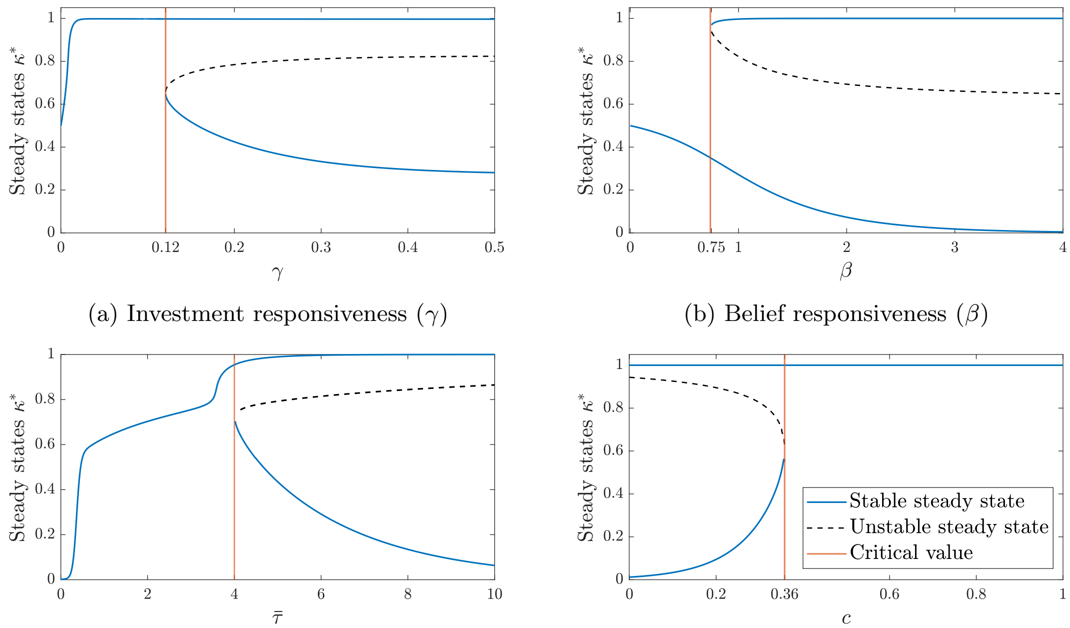
The existence of the high-carbon steady state is determined only by condition (17) on commitment. In particular, the policy-maker's commitment must be below the threshold value c^{***} (blue line in Figure 2), which, again, has an upper bound at $\frac{1}{2}$ and two other terms, μ_3 and a behavioural premium ν_{ch}. For small values of $\tilde{\lambda}_h$ ($\tilde{\lambda}_h < \frac{\chi_b-\chi_s}{2}$), the threshold c^{***} increases for higher values of belief responsiveness β (i.e. lower behavioural frictions), vulnerability to transition risks a and tax target $\bar{\tau}$, expanding the area of existence of the high-carbon steady state and thus increasing the likelihood of the high-carbon trap. The other parameters have instead ambiguous effects.

By combining the conditions presented in Proposition 3, we can characterise how the policy-maker's behaviour influences the long-term behaviour of the model under behavioural frictions, as illustrated in Figure 2. Some portions of the space are qualitatively similar to the ones identified in the neoclassical limit case (Figure 1): an ambitious and highly committed policy-maker drives the economy towards decarbonisation (upper-right quadrant); lack of ambition and commitment causes instead the transition to fail (lower-left quadrant); ambitious policy targets announced by an uncommitted policy-maker generate multiple steady states (mid-right quadrant). However, the presence of behavioural frictions leads to the emergence of two additional long-run behaviours. First, a committed policy-maker consistently meeting a non-ambitious tax target entails the existence of one or multiple mid-carbon steady states (upper-left quadrant). In this area, the equilibrium low-carbon capital share is lower than $1 - \lambda_l$ but higher than $\chi_s + \lambda_h$, suggesting less carbon-intensive steady states than those in the same quadrant of Figure 1. The reason is that, although under low tax targets high-carbon capital is cheaper than low-carbon capital, a portion of firms decide to invest in the more expensive low-carbon technology, due to particular environment-friendly preferences or to limited awareness of technology cost differences. Therefore, under a non ambitious but highly committed policy-maker, the presence of behavioural frictions is actually positive for the low-carbon transition. Second, non credible but ambitious tax announcements (bottom right corner) have worse effects on the low-carbon transition under behavioural frictions than in the neoclassical limit benchmark case. In fact, in the presence of behavioural frictions, the high-carbon sector never disappears entirely. As a result, transition risks remain positive and become considerable if policy targets are too ambitious, driving the actual policy away from targets, losing credibility and preventing the low-carbon steady state to exist.

3.3 High-carbon trap drivers

Figure 3 offers more details on the role of key parameters (γ, β, $\bar{\tau}$ and c) in determining the number and nature of the system steady states. We develop a numerical example based on our wider model calibration (see section 4), where we assume weak mitigation commitment by the policy-maker (i.e. $c = 0.3$). The blue solid lines represent stable steady states, while the black dashed lines are unstable intermediate steady states. The vertical red lines correspond to the critical value where the bifurcation occurs.

Figure 3a illustrates the impact of investment responsiveness γ on the position of steady states κ^*. For low values of γ, the model exhibits a unique stable equilibrium steeply increasing in the investment responsiveness. As γ passes the threshold indicated by the vertical red line ($\gamma \approx 0.12$ in this numerical example), a new stable high-carbon equilibrium emerges, together with an intermediate unstable one. That is, if investors are sufficiently responsive to expected cost differentials, the system could fall into either of the two steady states, depending on the initial conditions. If the economy is already sufficiently decarbonised (high κ), then the policy announcement can push low-carbon investment strongly enough to ensure a full transition before the policy-maker can lose its credibility due to weak commitment. Otherwise, despite a potential initial spur in low-carbon investments, the loss of credibility of the policy-maker eventually leads to a failure of the decarbonisation process. The relative sizes of the steady states' basins of attraction also move, in favour of the high-carbon one, as investment responsiveness γ increases.



(a) Investment responsiveness (γ)

(b) Belief responsiveness (β)

(c) Tax target ($\bar{\tau}$)

(d) Commitment level (c)

Figure 3: Bifurcation diagrams. Default parameter values are $\bar{\tau} = 6$, $c = 0.3$, $\gamma = 1$ and $\beta = 1$.

Figure 3b focuses on the impact of belief responsiveness β on long-term steady states of κ^*. It shows that for low values of belief responsiveness, the steady state is unique and decreasing in β, as a higher belief responsiveness leads weak commitment to be more punished by firms. As β passes a certain threshold ($\beta \approx 0.75$), two additional steady states emerge, one of which is low-carbon and stable. We thus confirm the importance of belief responsiveness in determining the nature of the system steady states, already pointed out in section 3.2 where higher β moves the green line in Figure 2 down and the blue line up, expanding the area where the low- and high-carbon steady states coexist. Further, in the area to the right of the critical value β^* in Figure 3b, an increase in belief responsiveness widens the low-carbon basin of attraction, but also leads to worse high-carbon steady states and better low-carbon steady states, i.e. low-carbon steady states closer to $\kappa^* = 1$.

Figure 3c shows the evolution of the long-term steady states as the tax target varies. As $\bar{\tau}$ grows, the unique steady state κ^* increases and the system approaches full decarbonisation. Under the current calibration, the nonlinear behaviour of κ^* with respect to the tax target is due to the fulfilment of conditions (16). When $\bar{\tau} > \bar{\tau}^{**}$ all believers fully invest in low-carbon capital but the overall share of believers grows slowly, as policy intensity is not sufficient to satisfy the second condition in (16); this corresponds to the first slowdown in Figure 3c. The second steep increase in κ^* is due to the fulfilment of such second condition in (16), which boosts the switch of sceptic firms to believers, making κ^* more sensitive to the policy strength. Finally, as the announced tax target passes the critical value indicated by the vertical red line ($\bar{\tau} \approx 4$ in our numerical example), two additional steady states emerge, one stable and one unstable. The resulting long-term low-capital share thus fundamentally depends on the basins of attraction of such steady states.

Finally, Figure 3d illustrates the bifurcation diagram of c. When c is lower than a critical value ($c \approx 0.36$), two additional steady states emerge through a fold bifurcation. Once this threshold is passed by c, the two lower steady states disappear and κ_l^* remains the unique fixed point. The critical value of c fundamentally depends on the tax target level, as shown in Propositions 2 and 3. In this respect, an additional key feature of the equilibrium conditions, both in the neoclassical limit and behavioural frictions scenarios, is that for $\bar{\tau} \to \infty$, the threshold of c tends to $\frac{1}{2}$, meaning that for very ambitious policy objectives, the policy-maker must strictly prefer meeting them than reducing transition risks, in order to avoid the emergence of a bad equilibrium. Therefore, announcing a sufficiently high tax target is needed in order to achieve the low-carbon transition, but being too ambitious in policy objectives increases the required commitment to those objectives. In other words, ambitious climate policies must be credible, or a high-carbon trap might emerge, potentially leading the low-carbon capital share to lower levels than under a less ambitious tax target.

3.4 A safe threshold for the carbon price announcements

From the policy-maker's perspective, the conditions for ensuring the existence (and uniqueness) of the low-carbon steady state and avoiding the high-carbon trap might be hard to estimate and, therefore, to apply. In addition, if we consider commitment as an intrinsic characteristic of the policy-maker, the only policy choice concerns the appropriate target tax schedule, given a specific c. Hence, we present an additional, simpler rule identifying a sufficient condition for the uniqueness of the equilibrium.

Proposition 4. *Under behavioural frictions, a unique low-carbon steady state exists if conditions* (16) *are met and:*

$$\bar{\tau} < \frac{1}{\beta(1-c)}. \tag{18}$$

Proof. Proof of proposition 4 is shown in Appendix B.4.

Proposition 4 identifies a sufficient condition: it states that the lower the policy-maker's commitment and the higher firms' belief responsiveness β, the less ambitious the policy announcement can be to guarantee the uniqueness of the low-carbon steady state.[15] The tax target implied in Proposition 4 can be interpreted as a safe threshold, below which the equilibrium is unique and able to generate a smooth transition, provided that it satisfies condition 16. The policy-maker could

[15] The same condition can be interpreted as an additional minimum commitment threshold, besides $c^{*}*$ in condition (16). This reads, given a certain announced tax rate, as follows: $c > 1 - \frac{1}{\bar{\tau}\beta}$.

therefore consider setting a tax slightly below the safe threshold to achieve full decarbonisation without risking the high-carbon credibility trap.

4 Calibration

We now remove the simplifying assumptions of the reduced version of the model used in section 3 and we calibrate and initialise the full model to European data. We use quarterly time steps, investigating the period 2020-2100 (with a total span of 320 time periods). Our baseline configuration is presented in Table 1. For all parameter values we deem as uncertain, we provide a sensitivity analysis in Appendix C.

4.1 Production

We study the transition to low-carbon capital in a growing economy where the growth rate of output, g_Y, is kept constant and equal to 0.5% (quarterly), corresponding to a yearly growth rate of approximately 2% (cfr. Lera and Sornette, 2017; van der Ploeg and Rezai, 2020; Gomme and Rupert, 2007). The quarterly depreciation rate, δ, is calibrated at 1.77%, consistent with Gomme et al. (2011) and corresponding to an annual depreciation of approximately 7.27% per period.

Firms' planning horizon, R, is set to 30 years, i.e. 120 quarters, corresponding to the average technical lifetimes of power plants (IEA, 2020c).[16] The yearly discount rate is set to 7%, as in IEA (2020c), corresponding to a quarterly discount rate ρ of approximately 1.7%. The implied quarterly discount factor D is 0.98.

We rely on data from the power sector to define the initial share of low-carbon capital κ_0, which is set to 0.2. In particular, we consider EU solar and wind installed electrical capacity in 2020 weighted by their capacity factors, i.e. 22289.32 and 64291.05 MW, respectively (Eurostat, 2021; IRENA, 2021).[17]. The total installed capacity adjusted by capacity factors includes, besides wind and solar, also combustible fuels, hydro, nuclear and other sources (geothermal and bioenergy-fired power), and amounts to 422142.91 MW (Eurostat, 2021).[18] Hence, the share of solar and wind capacity over total capacity approximately equals 20%. Under slightly different assumptions about

[16]Farfan and Breyer (2017) show that an average power plant technical lifetime of about 40 years for coal, 34 years for gas and 34 years for oil-fired power plants. Concerning renewable energy technologies, Tran and Smith (2018) considers lifetime for solar and wind plant ranging between 15 and 35 years.

[17]Solar and wind installed capacity amount to 138443 and 176985 MW, respectively (Eurostat, 2021) These values are then adjusted by solar and wind global weighted-average capacity factors, i.e. 16.1% and 36.3% (IRENA, 2021). The wind capacity factor corresponds to the average of onshore (36%) and offshore (40%) capacity factors weighted by their shares of EU wind installed capacity in 2020.

[18]The source-specific installed capacity are the following. The combustible fuels installed capacity is 388223 MW, while their average capacity factor is estimated at 47,9%, i.e. the average between the 2021 US capacity factor of coal (49.3%) and natural gas (45.76%), weighted by their shares (EIA, 2021). Hydro and nuclear installed capacity equal 150771 MW and 106008 MW, with a capacity factor of 46% and 80.3%, respectively. Other sources such as geothermal and bioenergy-fired power amount to 2171 MW with a capacity factor of 76.5%.

Table 1: Parameters and initial values in the central scenario

Parameter	Symbol	Quarterly value	Source
Production			
Output growth rate	g_Y	0.5%	Gomme and Rupert (2007); Lera and Sornette (2017)
Depreciation rate	δ	1.77%	Gomme et al. (2011)
Initial low-carbon capital share	κ_0	0.2	Eurostat (2021); IRENA (2021)
Low- to high-carbon production cost	$\frac{\theta_l}{\theta_h}$	1.36	IEA (2020c)
Beliefs and investment decisions			
Investment responsiveness	γ	1	Indirectly calibrated
Planning horizon	R	120	Farfan and Breyer (2017); Tran and Smith (2018)
Discount rate	ρ	1.7%	IEA (2020c)
Initial fitness measure differential	$U_{b,0}-U_{s,0}$	0.85	Refinitiv (2019)
Initial shares of belief types	$n_{b,0}$; $n_{s,0}$	0.3; 0.7	Refinitiv (2019)
Belief responsiveness	β	1	Ellen and Zwinkels (2010)
Belief-specific discount on targets	ϵ_b; ϵ_s	1; 0	Sensitivity analysis
Memory parameter	η	0.5	Hommes and Lustenhouwer (2019)
Policy			
Initial tax rate	τ_0	0.15	EEA (2021a,b)
Announced tax growth rate	$\bar{g}_\tau$	0.016	BEIS (2021); France Stratégie (2019); Riahi et al. (2017)
Transition risk index parameter	a	1	Indirectly calibrated

the energy sources to include, the years to consider and the likes, we obtain values for the initial low-carbon capital share ranging between 0.15 and 0.21. We thus run a sensitivity analysis on κ_0, and present the results in Appendix C.

The ratio of low- to high-carbon capital costs, $\frac{\theta_l}{\theta_h}$, is initialised to 1.36. This quantity proxies the relative convenience of capital investment in high-carbon technologies (see also Acemoglu et al., 2012; Lamperti et al., 2020; van der Ploeg and Rezai, 2020). We calibrate it by computing the average power generation costs of high-carbon (coal and gas) and low-carbon technologies (solar PV, wind onshore and wind offshore), excluding nuclear plants, in the EU countries at the end of 2019. In particular, we consider the average levelised cost of electricity (LCOE) for high- and low-carbon sources, which is, respectively, 80.6 and 109.6$/MWh,[19] indicating that high-carbon technologies exhibit a 36% advantage in cost-competitiveness.

4.2 Beliefs and investment decisions

Investment responsiveness, γ (cfr. equation 7), is indirectly calibrated to a value that allows the low-carbon transition - as proxied by the low-carbon share of total capital stock approaching 100% -

[19] Data are taken from the the Levelised Cost of Electricity Calculator of the International Energy Agency (IEA, 2020b). As we are interested in LCOEs under no carbon price, we first set the carbon price to zero and then compute the average LCOE for EU27. The geographic coverage of the calculator varies across technologies. Hence, not all European countries are included and the ones included vary across energy sources.

to occur by year 2100 in the benchmark scenario, i.e. with full commitment to climate objectives.[20] After running a battery of experiments, we set it to 1 and provide a sensitivity analysis in Appendix C. Our choice is motivated by the willingness to use a benchmark scenario wherein the transition to low-carbon investments gradually happens. From a quantitative perspective, $\gamma = 1$ implies that, given the initial backwardness of low-carbon technologies and assuming no climate policy, the share of low-carbon investment is lower than 1%. Further, this choice allows capturing some of the observed inertia in investment rebalancing processes (see also Waisman et al., 2012; Bilias et al., 2010; Vogt-Schilb et al., 2012).

We initialise the difference between beliefs' fitness measures to $U_{b,0} - U_{s,0} = 0.85$, so that the resulting initial share of believers equals $n_0 = 0.3$, corresponding to the proportion of participants to the 2019 Refinitiv Carbon Market Survey (Refinitiv, 2019) expressing trust in the policy-maker's announced strategy regarding the Market Stability Reserve of the EU Emission Trading Scheme.[21]

Estimates of the intensity of choice governing switching mechanisms between various expectation rules (β, see equation 4) have been at the core of intense debate. Though high values of such parameter predict strong responsiveness of economic agents towards more accurate expectations, several studies conclude in favour of relatively low values or even not significantly different from zero for β, especially when underlying data comes from financial markets (Boswijk et al., 2007; Kukacka and Barunik, 2017; Lamperti, 2018). For example, Chiarella et al. (2014) empirically assesses heterogeneous expectations in asset pricing, using a maximum likelihood approach on S&P500 data to estimate a structural model and finds the estimates between 0 and 1. Further, Ellen and Zwinkels (2010) estimates an oil price dynamics model with fundamentalist and chartist expectation rules and reports values for the intensity of choice around 1. However, it is important to stress that the calibration of the intensity of choice depends on the unit of measurement of the fitness measure and therefore is model specific. Hence, we select a value of 1, which is close to the majority of estimates available in the relevant literature (see also Hommes, 2021), but experiment with an ample range of alternative values in sections 5 and 6.

Following Hommes and Lustenhouwer (2019), we set the memory parameter, η, to 0.5, allowing agents to significantly update their evaluation of the heuristics when new information arises, but also to put considerable weight on the past.

Finally, our benchmark calibration assumes believers to fully trust the policy-maker's announcement, i.e. $\epsilon_b = 1$, and sceptics to discount the announced tax growth entirely, i.e. $\epsilon_s = 0$ and $E_s(g_\tau) = 0 \ \forall t$. This is the extreme case of sceptics expecting no carbon price increase at all. We perform extensive analyses on this parameter, whose exact value is - a priori - extremely difficult

[20] For additional information about indirect calibration of simulation models please see Fagiolo et al. (2019).

[21] Participants were asked whether they believed the permits' intake of the Market Stability Reserve (MSR) in case of excess allowances to be reduced at 12%, as planned at the time, or kept at the same level, i.e. 24%. Around 69% of respondents predicted the MSR intake rate to remain stable, with only 31% expecting the announced policy to actually be implemented. Although in this specific case the policy-maker announced a *less* stringent policy to be implemented in the future, we interpret this result as a proxy for the general trust in policy-makers' stated plans.

to identify without dedicated experiments.

4.3 Policy

The path of the announced climate policy is fully determined by the couple $\{\tau_0, \bar{g}_\tau\}$, which is composed by the initial tax rate imposed upon high-carbon capital and its announced growth rate.

These two policy parameters are calibrated following distinct strategies. In order to set the initial tax τ_0, we convert the carbon price per ton of CO2 to a tax on production costs. In 2020, the average allowance price in the EU-ETS was approximately 28\$/tCO2 (EEA, 2021a), increasing to 45\$/tCO2 in 2021. With respect to electricity generation, data on the EU emission intensity in 2020 range between 0.0002 tCO2/KWh (IEA, 2020a) and 0.0003 tCO2/KWh (EEA, 2021b). Hence, the carbon cost per kWh of electricity generated from fossil fuels varies between 0.006 and 0.014 \$, corresponding to a tax between 7.4-17% on the high-carbon average LCOE (0.0806 \$/KWh). We thus set $\tau_0 = 0.15$[22] and perform a sensitivity analysis across the above mentioned range of values in Appendix C, showing that the transition dynamics are qualitatively similar across such values.

The tax growth rate $\bar{g}_\tau$ is calibrated to 0.016 (equivalent to 1.6% quarterly and 6.6% annually) such that the projected carbon prices are aligned with those resulting from IPCC scenarios and those employed by the British and French governments. BEIS (2021) estimates the 2050 carbon 'value' to be within the range of 189-568£/tCO2e, with a central value of 378£/tCO2e, while France Stratégie (2019) proposes a shadow price of carbon between 600 and 900€/tCO2e for 2050. Under our model configuration, the carbon price reaches a value of approximately 314.34\$/tCO2 at the middle of the century. This calibration is also reasonably close to the average growth rate of the carbon price from mitigation pathway scenarios to 2100 taken from ENGAGE, an inter-model comparison project involving sixteen Integrated Assessment Models (IAMs) to design cost-effective pathways meeting the objectives of the Paris Agreement.[23] The average quarterly growth rate of the carbon price in the scenarios compatible with a 2°C temperature constrain is approximately between 2.4% and 2.9% (9.95 and 12.11% annually), depending on whether temperature overshoot is allowed.[24] The carbon price growth rates suggested by cost-efficient IAMs can be considered unrealistically high and driven by the implicit acceptance of a second-best scenario where the policy-maker will not accept large immediate price jumps in the short-term Gollier (2022). In the context of our model, what the optimal price is however less relevant, as we are primarily interested in capturing the actual 'political' projections communicated to the public opinion, however they are chosen.

Vulnerability to transition risks a (see equation 10) is set to 1, such that the transition risk

[22]Under the assumption of 0.00025 tCO2/KWh emission intensity, this initial tax corresponds to an initial carbon price approximately equal to 48.36\$/tCO2.

[23]All data about the scenarios we use can be retrieved at `https://data.ene.iiasa.ac.at/engage/#/workspaces`; see Riahi et al. (2021).

[24]More precisely, our set of scenarios assumes the pursue of Nationally Determined Contributions (NDC) until 2030, followed by the imposition a carbon budget between 1000 and 2000 GtCO2 IPCC (see also 2021).

index (i) is a smooth, increasing and concave function of the low-carbon share of capital and the tax target; (ii) is relatively low in 2020 ($\pi_0 \approx 0.11$); and (iii) would need a sharp increase of current policy stringency to rise above its mid-point value ($\pi_0 \approx 0.5$ for $\bar{\tau} \approx 1.2$, equivalent to a carbon price of 386.88 \$/tCO2), proxying a European economy relatively insensible to small variations of carbon prices (Metcalf and Stock, 2020) but potentially vulnerable to large shocks (Känzig, 2021).[25][26]Appendix C illustrates how the low-carbon capital share and the believers' share vary with a.

Finally, being a particularly key but ineffable value to calibrate, we explore the implications of the policy-maker's commitment level c (see equation 11) across its whole range of potential values $[0,1]$. However, we set two reference values for the parameter in sections 5 and 6: $c = 1$ represents our fully-committed policy-maker; $c = 0.3$, a value below the threshold causing the emergence of multiple steady states (see Figure 3), represents instead our weakly-committed policy-maker.

5 The transition dynamics under full commitment

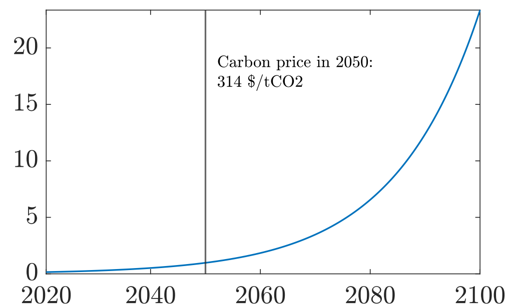
By simulating the calibrated model from 2020 to the end of the century, we first study how the economic system behaves when the policy-maker does not deviate from its climate policy targets ($c = 1 \ \forall \ t$). The full commitment scenario, whereby announced objectives are always achieved, provides a benchmark against which we are able to assess the transition dynamics under different parameter configurations. In what follows, we first illustrate the evolution over time of a set of key variables and then move to studying how the transition dynamics responds to varying degrees of behavioural frictions and opinion polarisation. We will later explore the case of low climate mitigation commitment in section 6.

5.1 The transition path

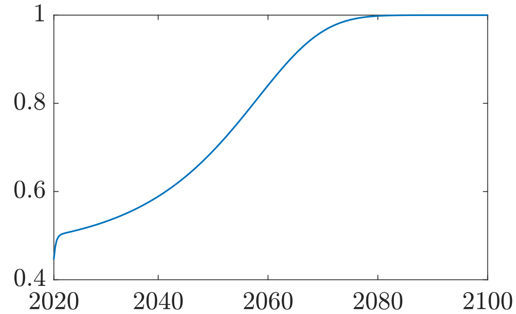
Figure 4 presents the evolution of a selection of key variables up to 2100. In this scenario, the carbon tax effectively imposed grows exponentially at the announced growth rate $\bar{g}_\tau$ (Figure 4a). As the policy-maker consistently meets its policy targets, the share of firms believing in the announced carbon price trajectory progressively increases (Figure 4b). The speed at which the policy-maker gains credibility depends both on the belief responsiveness, as measured by β, and on the relative accuracy of the two expectation rules (equation 4). Since the tax target grows gradually, the difference between believers' and sceptics' prediction accuracy is initially small. Therefore, at first firms are unable to precisely assess the policy-maker's credibility and the increase in the share of

[25]The impact of the parameter a on the implemented tax is given by $\frac{\partial \tau}{\partial a} = \frac{-(1-c)\bar{\tau}^2(1-\kappa)}{(1+a(1-\kappa)\bar{\tau})^2}$.

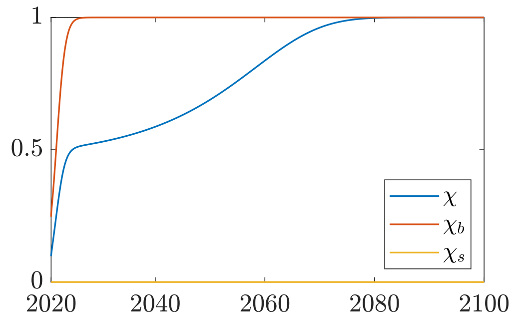
[26]As shown in Känzig (2021), the carbon policy surprise series employed is characterised by quite large variations in the carbon price, indicating large policy shocks.



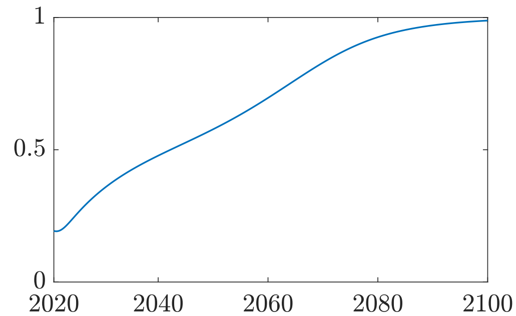
(a) Tax announced ($\bar{\tau}$)



(b) Share of believers (n)



(c) Low-carbon investment share (χ)



(d) Low-carbon share of capital (κ)

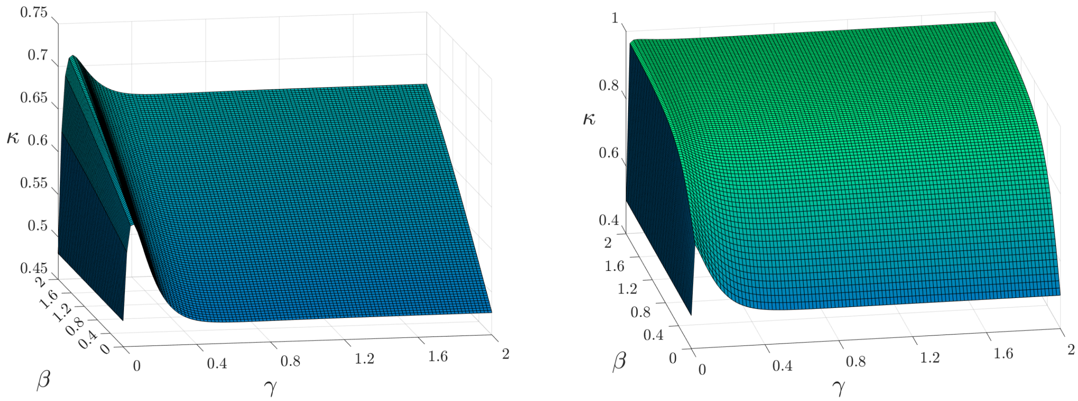
Figure 4: Evolution over time of selected variables under full commitment ($c = 1$).

believers is relatively slow. The shift later accelerates when it becomes clearer that believers are consistently correct in their expectations.

As illustrated in equation (8), the evolution of the aggregate clean investment share χ is determined by both the belief dynamics and the technology choices of each belief type χ_j (Figure 4c). While sceptics expect no tax increase and thus have a constant low-carbon investment share close to zero, believers progressively expand their clean investment share χ_b, which converges close to 1 around 2030. The resulting aggregate investment increases quite fast in the first decade, mainly thanks to believers rapidly increasing their clean investment share. After 2030, the aggregate clean investment share keeps growing, although at a lower rate, thanks to the increase of the share of believers in the population. It eventually reaches a value close to one around 2070.

Low-carbon investment drives the dynamics of the clean capital share κ, as in equation (9). Figure 4d illustrates its evolution through the century: with a policy-maker fully committed to climate objectives, more than 60% of capital is low-carbon by 2060 and the transition is fully achieved by the end of the century.[27]

[27]Note that, as detailed in section 3, under finite belief and investment intensities of choice, an infinitesimal portion of high-carbon capital continues to exist even in the low-carbon steady state.



(a) 2050

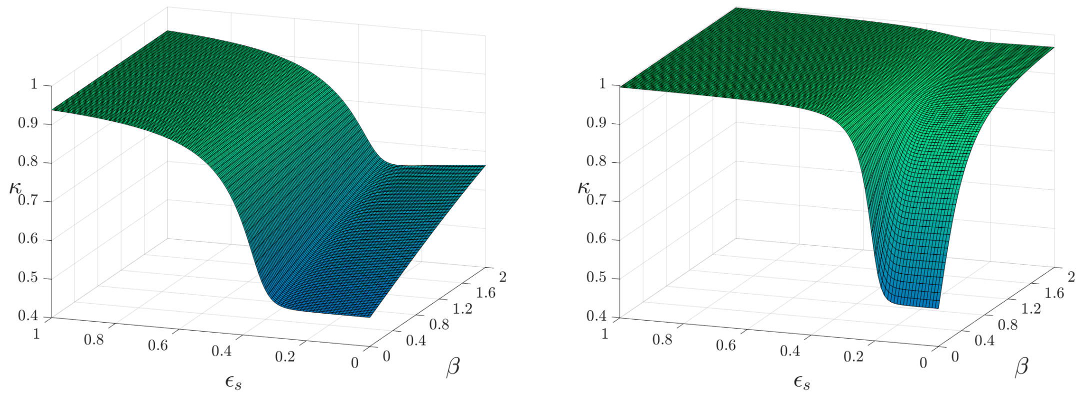
(b) 2080

Figure 5: Low-carbon capital share κ as a function of belief responsiveness β and investment responsiveness γ, under $c = 1$, in (a) 2050 and (b) 2080. All other parameters at their default value (Table 1).

5.2 Behavioural frictions and opinion polarisation

The transition pathway does not depend solely on policy targets and policy-maker's behaviour, but also on how firms respond to them. Even under full commitment, the transition might be hampered if firms fail to internalise long-term policy objectives and/or are reluctant to modify their investment choices. In what follows, we present snapshots of the clean capital share κ at selected years under different behavioural configurations.

Figure 5 shows clean capital share κ in 2050 (panel (a)) and 2080 (panel (b)) for various degrees of belief and investment responsiveness (β and γ, respectively), which, as discussed in section 2, can be thought of as inversely related to behavioural frictions. A fully committed policy-maker is able to eventually achieve (almost) full decarbonisation under most behavioural configurations, except when firms are entirely unresponsive to its policy choices (i.e. if β and/or γ equal zero). Excluding these corner parameter values, the speed of transition greatly varies depending on behavioural frictions. A lower responsiveness of firms to prediction errors (lower β) is undesirable, as it hampers clean investment and delays decarbonisation. The impact of investment responsiveness, instead, is mixed. *Ceteris paribus*, when firms are more responsive to expected cost differences (higher γ), the biases of both believers and sceptics (towards clean and dirty choices, respectively) are amplified, leading them to allocate a higher investment share to their preferred technology. As a result, the impact of investment responsiveness on the transition pace varies with belief dynamics. In the first decades, an increase in γ first accelerates the transition, because of the positive impact on believers' clean investment, but later, as γ crosses a certain threshold, it disproportionately lowers sceptics'



(a) 2050

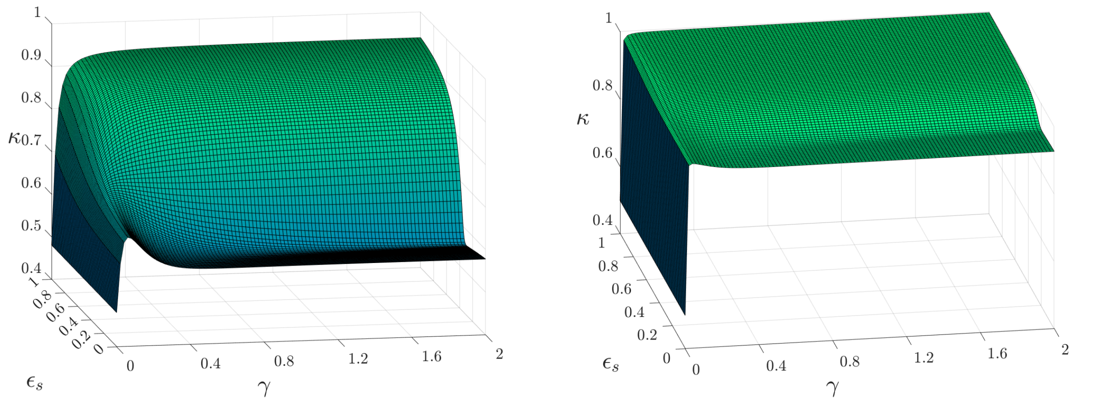
(b) 2080

Figure 6: Low-carbon capital share κ as a function of sceptics' discounting of the tax target growth rate ϵ_s and belief responsiveness β, under $c = 1$, in (a) 2050 and (b) 2080. All other parameters at their default value (Table 1).

clean investment, hampering the transition. Over time, however, for sufficiently high values of belief responsiveness, sceptics rapidly disappear from the population and higher investment responsiveness does not slow down the transition.

Figure 6 explores how the clean capital share varies with belief responsiveness and opinion polarisation. As mentioned in section 2.1, assuming a fixed ϵ_b, we can use the degree of trust of sceptics in the announced tax target growth rate, $\epsilon_s \in [0, 1]$, as our proxy for opinion polarisation. Through the entire transition, the absence of opinion polarisation ($\epsilon_s = \epsilon_b = 1$) leads to the highest low-carbon capital shares – above 87% in 2050 and close to 100% in 2080, regardless of β. When all firms believe in policy announcements, the transition dynamics depends solely on the tax schedule targeted by the policy-maker and belief responsiveness has no impact whatsoever. For low values of β, the low-carbon share of capital increases monotonously in ϵ_s: as firms are split almost equally between the two expectation rules, the more these converge towards believing the policy announcements, the higher the low-carbon share of capital reached in a certain period.

For higher values of β, opinion polarisation has a slightly non-monotonous impact on the transition dynamics. In particular, as belief responsiveness crosses a certain threshold, a very high polarisation between beliefs (i.e. $\epsilon_s = 0$), leads to a faster low-carbon transition than intermediate values of ϵ_s. The reason lies in the belief switching mechanism. When sceptics expect the tax to be constant over time ($\epsilon_s = 0$) but the policy-maker is fully committed to its policy targets, sceptics' predictions end up being inaccurate soon. Thus, their prediction errors lead firms to switch to the believer expectation rule. Also, the higher β the sooner this belief switch occurs, as firms react rapidly to prediction errors. On the other hand, sceptics mildly discounting the tax target growth



(a) 2050 (b) 2080

Figure 7: Low-carbon capital share κ as a function of sceptics' discounting of the tax target growth rate ϵ_s and investment responsiveness γ, under $c = 1$, in (a) 2050 and (b) 2080. All other parameters at their default value (Table 1).

rate (values of ϵ_s closer to 1), lead sceptics' prediction errors to be not large enough to induce a rapid switch in beliefs, especially in the first part of the simulation, when the tax target is still low. Over time, this non-monotonicity is reduced because, eventually, sceptics disappear from the population of firms, leading to high values of clean capital share κ.

Finally, keeping belief responsiveness fixed at its default calibration ($\beta = 1$), we confirm the non-monotonous impact of investment responsiveness and opinion polarisation on the pace of decarbonisation (Figure 7). More specifically, we find that, when firms are highly responsive to expected cost differences (high γ), decreasing opinion polarisation (varying ϵ_s from zero to approximately 0.4) slightly decreases the transition speed. This non-monotonicity is hardly visible because it results from two balancing forces of roughly equal force. On the one hand, decreasing opinion polarisation under full commitment slows down the disappearance of sceptics from the population of firms. On the other hand, a higher ϵ_s diminishes the sceptics' bias towards high-carbon technologies. As ϵ_s passes a certain threshold, especially in 2050, the second effect strongly grows in importance, speeding up the transition. Further in time (panel (b)), if the policy-maker is committed to its policy targets, investment responsiveness and opinion polarisation, while still having a non monotonous impact, do not significantly hamper the transition.

6 The credible commitment problem

We now investigate transition patterns with varying policy-maker's mitigation commitment levels, as measured by parameter $c \in [0, 1]$. When the policy-maker is not fully committed to meeting

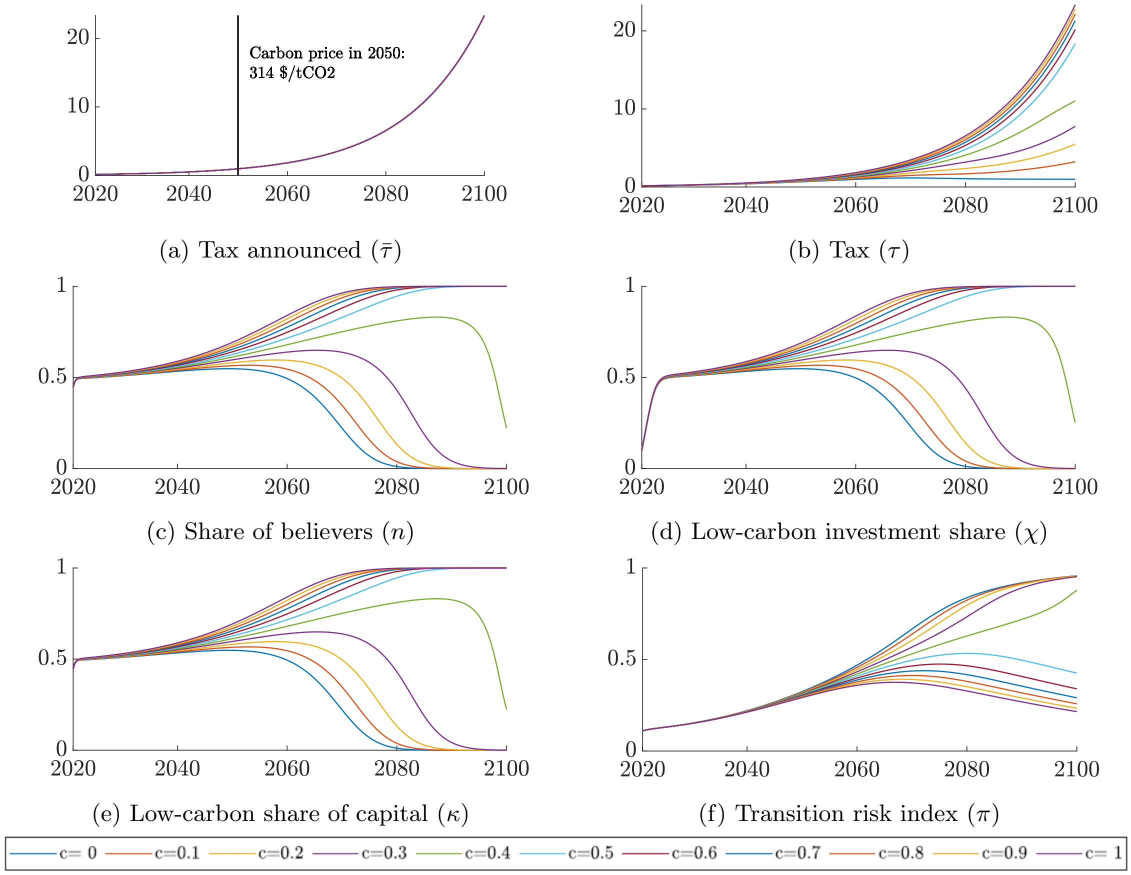


Figure 8: Evolution over time of selected variables under various levels of commitment (c).

its previously announced policy targets ($c < 1$), it might partly deviate from them in response to the perceived transition risks, as in equation (11). As shown in section 3, this departure from policy targets lies behind the emergence of multiple equilibria and might lead the economy into a high-carbon trap. In what follows, we illustrate the transition paths towards the various steady states of the model and the role played by behavioural frictions and opinion polarisation in the transition dynamics under weak commitment.

6.1 The transition path

Figure 8 presents the evolution over time of a set of key variables. Panel (a) shows the tax target schedule, which is independent of commitment c. The tax actually implemented, instead, is flatter for lower levels of commitment (panel (b)), because of the policy-maker's attempt to reduce higher

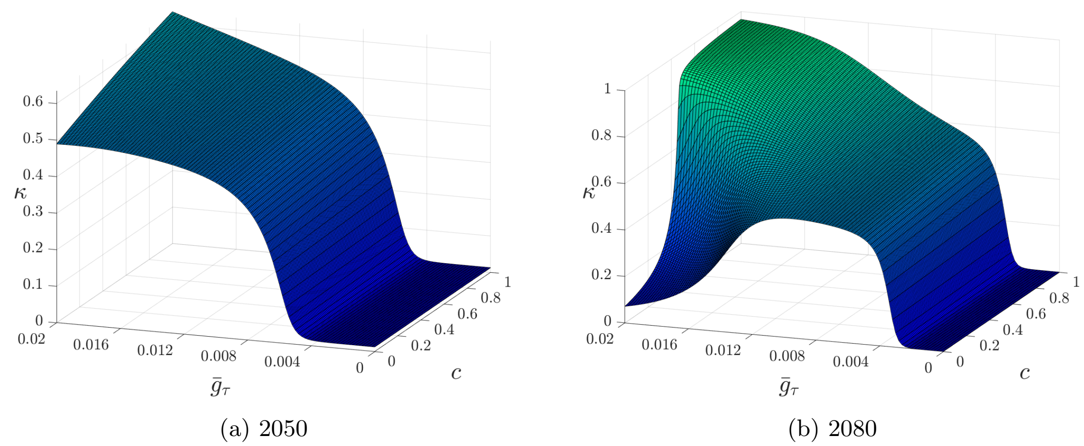


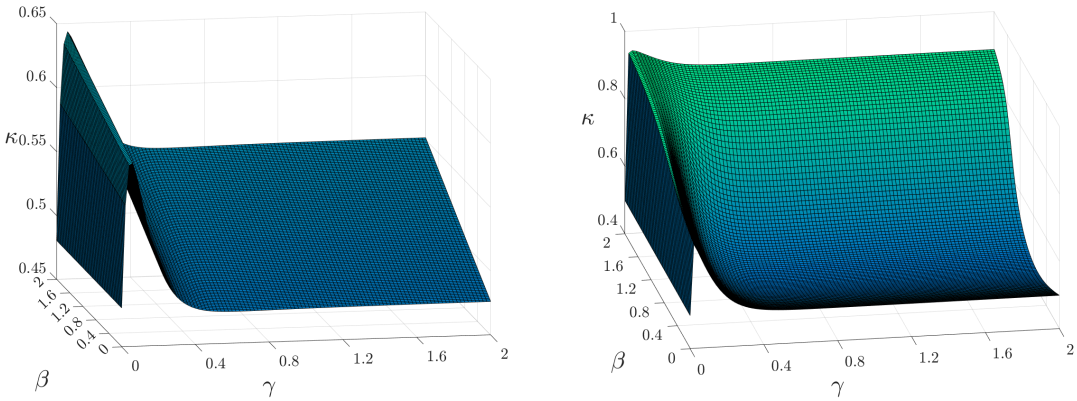
Figure 9: Low-carbon capital share κ as a function of the tax target growth rate $\bar{g}_\tau$ and commitment c, in (a) 2050 and (b) 2080. All other parameters at their default value (Table 1).

transition risks. As a result of mild policy stringency and default on its pledges, the non-committed policy-maker pays a cost in terms of credibility loss, reflected in the slower increase of believers' share n (panel (c)). This ultimately leads the low-carbon transition to slow down or, for low enough values of c, to fail entirely (panel (e)). In this case the system goes back to a dynamics characterised by a majority of sceptics ($n \approx 0$) and low clean investment and capital shares ($\chi \approx 0$; $\kappa \approx 0$).

The transition failure emerges only quite late in time: at first, even a weakly committed policy-maker cannot depart too much from policy targets, as these are initially mild and the associated transition risks low. In the first periods, therefore, the policy-maker's credibility is at worst more or less constant. As the tax target and transition risks grow over time, firms recognise the weakly committed policy-maker and eventually lose trust in its announcements, up to the point where low-carbon investment is significantly reduced. As this tipping point is reached, the transition dynamics reverses and κ converges to the sceptics' clean investment share, χ_s.

Furthermore, transition risks end up being significantly larger under weak rather than stronger commitment levels, because the high-carbon share of the economy affected by the tax is larger. Hence, the weakly committed policy-maker cannot really escape transition risks, but only postpone (and amplify) them to the future. The result is a higher cost of policy action, which further reduces policy-makers' options, up to the point where the low-carbon transition fails.

Figure 9 shows how the low credibility trap illustrated in section 3 emerges over time under an ambitious but uncommitted policy-maker. For several decades the low-carbon capital share is increasing in the tax target but is almost independent from the policy-maker's commitment, as transition risks are perceived to be mild and even the weakly committed policy-maker does not depart excessively from the announced targets (panel (a)). However, unmet ambitious objectives



(a) 2050 (b) 2080

Figure 10: Low-carbon capital share κ as a function of belief responsiveness β and investment responsiveness γ, under $c = 0.3$, in (a) 2050 and (b) 2080. All other parameters at their default value (Table 1).

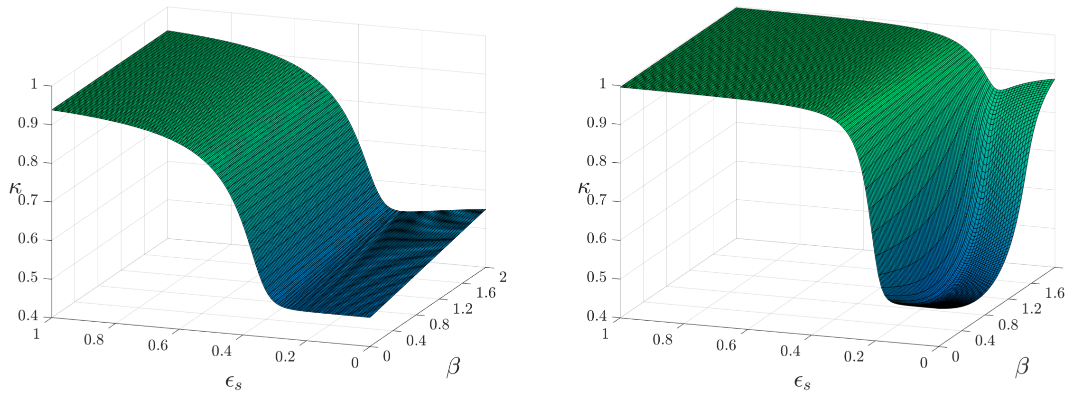
ultimately generate a credibility loss, leading to an increase in the population of sceptics and a decrease in the low-carbon capital share, ultimately causing a transition failure (panel (b)).

6.2 Behavioural frictions and opinion polarisation

The policy-maker's behaviour is a key determinant of the transition success or failure. However, as already shown in section 5.2, firms' response to policy choices is equally important for rapidly achieving decarbonisation. In what follows, we explore the impact of behavioural factors on the transition dynamics, under weak commitment ($c = 0.3$).

Figure 10 shows snapshots of the clean capital share κ at different points of time for various levels of belief and investment responsiveness. In 2050, the low-carbon capital share presents a similar relationship with β and γ to that under full commitment (Figure 5), although at lower levels of κ. The picture is instead different a few decades later (panel (b)). Indeed, belief responsiveness has a non monotonous effect on the clean capital share. Values of β slightly larger than 0 hamper the transition as firms realise the government is not keeping its word. As β crosses a certain threshold, the transition, especially in the early decades, is faster and involves lower transition risks faced by the policy-maker, who can thus implement a tax sufficiently close to the target.

Figure 11 illustrates how κ evolves under various levels of belief responsiveness β and opinion polarisation, proxied by ϵ_s. For sufficiently high values of belief responsiveness, strong opinion polarisation (i.e. low ϵ_s) accelerates the transition, especially in the early decades, with respect to milder polarisation. While the same effect takes place under full commitment (Figure 6), under

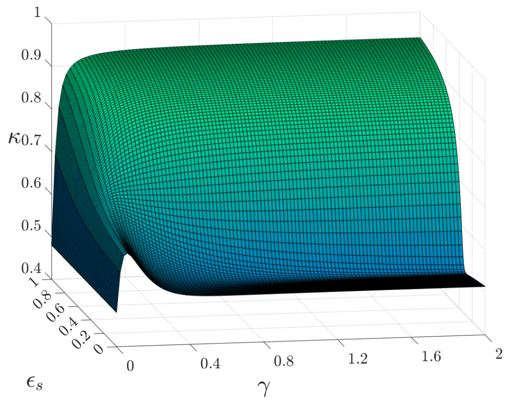


(a) 2050

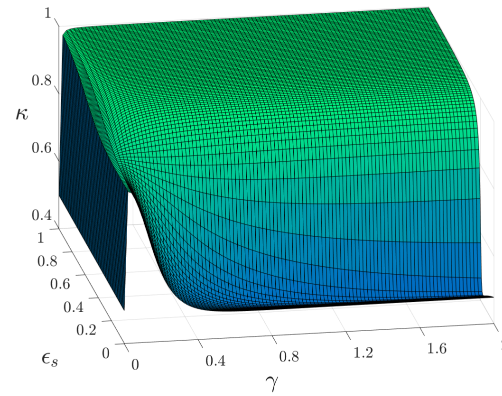
(b) 2080

Figure 11: Low-carbon capital share κ as a function of sceptics' discounting of the tax target growth rate ϵ_s and belief responsiveness β, under $c = 0.3$, in (a) 2050 and (b) 2080. All other parameters at their default value (Table 1).

weak policy-maker's commitment the non-monotonicity is much more pronounced. The reason is that, for low values of commitment c, a certain degree of scepticism produces even more accurate tax predictions than under full commitment, causing a delay in the disappearance of sceptics from the population of firms. Eventually, for high values of β, sceptics are proved wrong and the transition takes place, as illustrated in panel (b). For low belief responsiveness and high opinion polarisation, instead, the negative feedback loop between policy-maker's weak credibility and firms' investment choices emerges, pushing the economy into a high-carbon trap.



(a) 2050



(b) 2080

Figure 12: Low-carbon capital share κ as a function of sceptics' discounting of the tax target growth rate ϵ_s and investment responsiveness γ, under $c = 0.3$, in (a) 2050 and (b) 2080. All other parameters at their default value (Table 1).

Finally, Figure 12 shows the impact of investment responsiveness γ and opinion polarisation ϵ_s on the speed of decarbonisation. While in 2050 the figure is very similar to that under full commitment (see Figure 7), in 2080, under high investment responsiveness γ and high opinion polarisation (i.e. small ϵ_s), the weakly committed policy-maker forces the economy into a high-carbon trap. The reason is that both these behavioural factors increase sceptics' bias towards high-carbon technologies, depressing sceptics' low-carbon investment share, up to the point where the policy-maker fails to decarbonise the economy.

7 Conclusions

In this paper, we model and analyse the dynamic interaction between heterogeneous expectations, investment decisions and climate mitigation policy-making. We develop a novel modelling approach rooted in discrete choice theory, able to account for dynamic beliefs and policy uncertainty. We obtain four key results. First, a 'high-carbon credibility trap' - i.e. the convergence to a carbon-intensive steady state when an alternative low-carbon equilibrium is also present - might emerge when an ambitious plan is announced by a weakly committed policy-maker. This can trap the economic system in a vicious circle of carbon-intensive investment, increasing transition risks and policy-maker's credibility, eventually leading to a transition failure. Second, the presence of behavioural frictions - either in capital investment choices or in the assessment of the policy-maker's credibility - creates heterogeneous expectations and affects the conditions of existence of long-run system equilibria, making it harder to achieve full decarbonisation. However, higher responsiveness to the performance of belief/investment strategies makes firms' behaviour more volatile, increasing the likelihood to fall into a high-carbon trap. Third, even when the economic system is directed towards a low-carbon equilibrium (e.g. under a fully committed policy-maker), behavioural frictions affect the rapidity of the decarbonisation process in non-linear manners. Finally, belief polarisation can also have non-linear implications on decarbonisation, with higher belief polarisation being beneficial for the transition under certain circumstances.

Our work can benefit from a number of additional refinements. Our numerical model suffers from the scarce availability of systematic data concerning transition-related beliefs and expectations, making it hard to calibrate behavioural parameters (primarily γ and β). Consequently, the exact timing of our transition dynamics should mainly be interpreted in a qualitative manner, rather than a precise forecast. The complexity of the modelling framework could also be expanded, or directed towards additional research questions. For instance, we rely on an exogenous growth path to better focus on the investment allocation choice, but the transition dynamics is likely to have wider macroeconomic implications, suggesting additional insights could be obtained by making growth dynamics endogenous. Our commitment level c, now an intrinsic and immutable feature of a policy-maker, could also be made endogenous and variable in time, possibly jumping following electoral cycles. Another possible direction of research is to study financial, rather than capital, investment decisions, which would require incorporating a financial sector in the model. Finally, we abstract here from premature decommissioning, loss of capacity utilisation and costly capital reconversion ('stranding'), although including these dimensions, both in reality and in expectations, is likely to have implications on the overall transition dynamics (see for instance Cahen-Fourot et al., 2022; Campiglio et al., 2022)

Despite these limitations, our results offer several key insights for policy-making. We have shown how the direction and heterogeneity of expectations of future climate mitigation policies can significantly affect the dynamics of the climate policies themselves. It is thus absolutely key for the policy-maker to (i) be aware of what these expectations are and their distribution; and (ii) be able to orient them as desired. As mentioned above, the current availability of data on transition-related expectations, their drivers and their dynamic behaviour is very scarce. Public institutions should contribute to running surveys, experiments and empirical analysis that could, in combination, provide a more solid calibration basis in the future. The ability to orient expectations comes instead from credibility, itself a function of the past track-record in sticking to announced plans. In the context of climate policies, many jurisdictions have shown worrying signs of being unable to maintain their course for sufficiently extended periods of time. A wide societal debate on the most appropriate institutional configuration to achieve long-termist and credible policies is urgently needed. Finally, our results warn policy-makers of the risks of having both too little and too much ambitions in their policy announcements, as excessively ambitious plans combined with less than full commitment to respect them might backfire and produce worse outcomes than a less ambitious policy announcement.

References

Acemoglu, D., P. Aghion, L. Bursztyn, and D. Hemous (2012): "The Environment and Directed Technical Change," *American Economic Review*, 102, 131–166.

Annicchiarico, B., F. Di Dio, and F. Diluiso (2022): "Climate Actions, Market Beliefs and Monetary Policy," CEIS Working Paper 535, Centre for Economic and International Studies, Rome.

Assenza, T., P. Heemeijer, C. H. Hommes, and D. Massaro (2021): "Managing self-organization of expectations through monetary policy: A macro experiment," *Journal of Monetary Economics*, 117, 170–186.

Barradale, M. J. (2014): "Investment under uncertain climate policy: A practitioners' perspective on carbon risk," *Energy Policy*, 69, 520–535.

Barro, R. J. and D. B. Gordon (1983): "A Positive Theory of Monetary Policy in a Natural Rate Model," *Journal of Political Economy*, 91, 589–610.

Battiston, S., I. Monasterolo, K. Riahi, and B. J. van Ruijven (2021): "Accounting for finance is key for climate mitigation pathways," *Science*, 372, 918–920.

BEIS (2021): "Valuation of greenhouse gas emissions: for policy appraisal and evaluation," Policy paper, UK Department for for Business, Energy, & Industrial Strategy, London.

Berestycki, C., S. Carattini, A. Dechezleprêtre, and T. Kruse (2022): "Measuring and assessing the effects of climate policy uncertainty," Tech. rep., OECD, Paris.

Bilias, Y., D. Georgarakos, and M. Haliassos (2010): "Portfolio Inertia and Stock Market Fluctuations," *Journal of Money, Credit and Banking*, 42, 715–742.

Bordalo, P., N. Gennaioli, and A. Shleifer (2022): "Overreaction and Diagnostic Expectations in Macroeconomics," *The Journal of Economic Perspectives*, 36, 223–244.

Bosetti, V. and D. G. Victor (2011): "Politics and Economics of Second-Best Regulation of Greenhouse Gases: The Importance of Regulatory Credibility," *The Energy Journal*, 32, 1–24.

Boswijk, H. P., C. H. Hommes, and S. Manzan (2007): "Behavioral heterogeneity in stock prices," *Journal of Economic Dynamics and Control*, 31, 1938–1970.

Brock, W. A. and C. H. Hommes (1997): "A Rational Route to Randomness," *Econometrica*, 65, 1059–1096.

——— (1998): "Heterogeneous beliefs and routes to chaos in a simple asset pricing model," *Journal of Economic Dynamics and Control*, 22, 1235–1274.

Cafferata, A., M. J. Dávila-Fernández, and S. Sordi (2021): "Seeing what can(not) be seen: Confirmation bias, employment dynamics and climate change," *Journal of Economic Behavior & Organization*, 189, 567–586.

Cahen-Fourot, L., E. Campiglio, L. Daumas, M. G. Miess, and A. A. Yardley (2022): "Stranding Ahoy? Heterogeneous Transition Beliefs and Capital Investment Choices," .

Campiglio, E., S. Dietz, and F. Venmans (2022): "Optimal Climate Policy As If the Transition Matters," CESifo Working Paper 10139, CESifo, Munich.

Campiglio, E. and F. van der Ploeg (2022): "Macrofinancial Risks of the Transition to a Low-Carbon Economy," *Review of Environmental Economics and Policy*, 000–000.

Carattini, S., M. Carvalho, and S. Fankhauser (2018): "Overcoming public resistance to carbon taxes," *WIREs Climate Change*, 9, e531, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcc.531.

Chiarella, C., X.-Z. He, and R. C. J. Zwinkels (2014): "Heterogeneous expectations in asset pricing: Empirical evidence from the S&P500," *Journal of Economic Behavior & Organization*, 105, 1–16.

Comerford, D. and A. Spiganti (2022): "The Carbon Bubble: climate policy in a fire-sale model of deleveraging," *The Scandinavian Journal of Economics*, n/a, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/sjoe.12519.

Crowley, K. (2017): "Up and down with climate politics 2013–2016: the repeal of carbon pricing in Australia," *WIREs Climate Change*, 8, e458.

Daniel, K. D., R. B. Litterman, and G. Wagner (2019): "Declining CO2 price paths," *Proceedings of the National Academy of Sciences*, 116, 20886–20891.

De Grauwe, P. and C. Macchiarelli (2015): "Animal spirits and credit cycles," *Journal of Economic Dynamics and Control*, 59, 95–117.

Douenne, T. and A. Fabre (2020): "French attitudes on climate change, carbon taxation and other climate policies," *Ecological Economics*, 169, 106496.

Dunz, N., A. Naqvi, and I. Monasterolo (2021): "Climate sentiments, transition risk, and financial stability in a stock-flow consistent model," *Journal of Financial Stability*, 54, 100872.

Dávila-Fernández, M. J. and S. Sordi (2020): "Attitudes towards climate policies in a macrodynamic model of the economy," *Ecological Economics*, 169, 106319.

EEA (2021a): "National emissions reported to the UNFCCC and to the EU Greenhouse Gas Monitoring Mechanism," Tech. rep., European Environment Agency.

——— (2021b): "Trends and projections in the EU ETS in 2021 The EU Emissions Trading System in numbers," Tech. rep., European Environment Agency.

EIA (2021): "Electric Power Monthly - U.S. Energy Information Administration (EIA)," Tech. rep., Energy Information Administration.

Ellen, S. t. and R. C. J. Zwinkels (2010): "Oil price dynamics: A behavioral finance approach with heterogeneous agents," *Energy Economics*, 32, 1427–1434.

Engle, R. F., S. Giglio, B. Kelly, H. Lee, and J. Stroebel (2020): "Hedging Climate Change News," *The Review of Financial Studies*, 33, 1184–1216.

Eurostat (2021): "Electricity and heat statistics," Tech. rep., Eurostat, Luxembourg.

Evans, G. W. and S. Honkapohja (2006): "Monetary Policy, Expectations and Commitment," *The Scandinavian Journal of Economics*, 108, 15–38.

——— (2012): *Learning and Expectations in Macroeconomics*, Princeton University Press.

Fagiolo, G., M. Guerini, F. Lamperti, A. Moneta, and A. Roventini (2019): "Validation of Agent-Based Models in Economics and Finance," in *Computer Simulation Validation: Fundamental Concepts, Methodological Frameworks, and Philosophical Perspectives*, ed. by C. Beisbart and N. J. Saam, Cham: Springer International Publishing, Simulation Foundations, Methods and Applications, 763–787.

Fankhauser, S., S. M. Smith, M. Allen, K. Axelsson, T. Hale, C. Hepburn, J. M. Kendall, R. Khosla, J. Lezaun, E. Mitchell-Larson, M. Obersteiner, L. Rajamani, R. Rickaby, N. Seddon, and T. Wetzer (2022): "The meaning of net zero and how to get it right," *Nature Climate Change*, 12, 15–21.

Farfan, J. and C. Breyer (2017): "Aging of European power plant infrastructure as an opportunity to evolve towards sustainability," *International Journal of Hydrogen Energy*, 42, 18081–18091.

France Stratégie (2019): "The value for climate action. A shadow price of carbon for evaluation of investments and public policies," Report by the Commission chaired by Alain Quinet, France Stratégie.

Fried, S., K. Novan, and W. B. Peterman (2022): "Climate policy transition risk and the macroeconomy," *European Economic Review*, 147, 104174.

Fuss, S., J. Szolgayova, M. Obersteiner, and M. Gusti (2008): "Investment under market and climate policy uncertainty," *Applied Energy*, 85, 708–721.

Galanis, G., I. Kollias, I. Leventidis, and J. Lustenhouwer (2022a): "Generalizing Heuristic Switching Models," *Working Papers*, number: 0715 Publisher: University of Heidelberg, Department of Economics.

Galanis, G., G. Ricchiuti, and B. Tippet (2022b): "The Global Political Economy of a Green Transition," *Working Papers - Economics*, number: wp2022_22.rdf Publisher: Universita' degli Studi di Firenze, Dipartimento di Scienze per l'Economia e l'Impresa.

Gigerenzer, G. and H. Brighton (2009): "Homo Heuristicus: Why Biased Minds Make Better Inferences," *Topics in Cognitive Science*, 1, 107–143.

Gollier, C. (2022): "The cost-efficiency carbon pricing puzzle," TSE Working Papers 952, Toulouse School of Economics, number: 18-952 Publisher: Toulouse School of Economics (TSE).

Golosov, M., J. Hassler, P. Krusell, and A. Tsyvinski (2014): "Optimal taxes on fossil fuel in general equilibrium," *Econometrica*, 82, 41–88.

Gomme, P., B. Ravikumar, and P. Rupert (2011): "The return to capital and the business cycle," *Review of Economic Dynamics*, 14, 262–278.

Gomme, P. and P. Rupert (2007): "Theory, measurement and calibration of macroeconomic models," *Journal of Monetary Economics*, 54, 460–497.

Hale, T., T. Kuramochi, J. Lang, B. Mapes, S. Smith, R. Aiyer, R. Black, M. Boot, P. Chalkley, F. Hans, N. Hay, A. Hsu, N. Höhne, S. Mooldijk, and T. Walsh (2022): "Net Zero Tracker," Tech. rep., Energy and Climate Intelligence Unit, Data-Driven EnviroLab, NewClimate Institute, Oxford Net Zero, Oxford.

Helm, D., C. Hepburn, and R. Mash (2003): "Credible Carbon Policy," *Oxford Review of Economic Policy*, 19, 438–450.

Hommes, C. (2021): "Behavioral and Experimental Macroeconomics and Policy Analysis: A Complex Systems Approach," *Journal of Economic Literature*, 59, 149–219.

Hommes, C. and J. Lustenhouwer (2019): "Inflation targeting and liquidity traps under endogenous credibility," *Journal of Monetary Economics*, 107, 48–62.

Hommes, C. H. (2006): "Chapter 23 Heterogeneous Agent Models in Economics and Finance," in *Handbook of Computational Economics*, ed. by L. Tesfatsion and K. L. Judd, Elsevier, vol. 2, 1109–1186.

IEA (2020a): "Development of CO2 emission intensity of electricity generation in selected countries, 2000-2020 – Charts – Data & Statistics," Tech. rep., International Energy Agency.

——— (2020b): "Levelised Cost of Electricity Calculator," Tech. rep., International Energy Agency.

——— (2020c): "Techno-economic inputs – World Energy Model – Analysis," Tech. rep., International Energy Agency.

IPCC (2021): *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou, cambridge university press ed.

——— (2022): *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, Intergovernmental Panel on Climate Change.

IRENA (2021): "Renewable Power Generation Costs in 2020," Tech. rep., International Renewable Energy Agency.

Konc, T., S. Drews, I. Savin, and J. C. J. M. van den Bergh (2022): "Co-dynamics of climate policy stringency and public support," *Global Environmental Change*, 74, 102528.

Krueger, P., Z. Sautner, and L. T. Starks (2020): "The Importance of Climate Risks for Institutional Investors," *The Review of Financial Studies*, 33, 1067–1111, publisher: Oxford Academic.

Kukacka, J. and J. Barunik (2017): "Estimation of financial agent-based models with simulated maximum likelihood," *Journal of Economic Dynamics and Control*, 85, 21–45.

Kydland, F. E. and E. C. Prescott (1977): "Rules Rather than Discretion: The Inconsistency of Optimal Plans," *Journal of Political Economy*, 85, 473–491, publisher: The University of Chicago Press.

Känzig, D. R. (2021): "The Unequal Economic Consequences of Carbon Pricing," SSRN Scholarly Paper ID 3786030, Social Science Research Network, Rochester, NY.

LAMPERTI, F. (2018): "Empirical validation of simulated models through the GSL-div: an illustrative application," *Journal of Economic Interaction and Coordination*, 13, 143–171.

LAMPERTI, F., G. DOSI, M. NAPOLETANO, A. ROVENTINI, AND A. SAPIO (2020): "Climate change and green transitions in an agent-based integrated assessment model," *Technological Forecasting and Social Change*, 153, 119806.

LERA, S. AND D. SORNETTE (2017): "Secular Bipolar Growth Rate of the Real US GDP Per Capita: Implications for Understanding Past and Future Economic Growth," SSRN Scholarly Paper ID 2703882, Social Science Research Network, Rochester, NY.

MCCULLOCH, N., D. NATALINI, N. HOSSAIN, AND P. JUSTINO (2022): "An exploration of the association between fuel subsidies and fuel riots," *World Development*, 157, 105935.

MCFADDEN, D. (1974): "Conditional logit analysis of qualitative choice behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka, New York: Academic Press, 105–142.

MERCURE, J.-F. (2015): "An age structured demographic theory of technological change," *Journal of Evolutionary Economics*, 25, 787–820.

METCALF, G. E. AND J. H. STOCK (2020): "Measuring the Macroeconomic Impact of Carbon Taxes," *AEA Papers and Proceedings*, 110, 101–106.

NEMET, G. F., M. JAKOB, J. C. STECKEL, AND O. EDENHOFER (2017): "Addressing policy credibility problems for low-carbon investment," *Global Environmental Change*, 42, 47–57.

NOAILLY, J., L. NOWZOHOUR, AND M. VAN DEN HEUVEL (2021): "Heard the News? Environmental Policy and Clean Investments," CIES Research Paper 70, Graduate Institute of International and Development Studies, Centre for International Environmental Studies, Geneva.

NORDENG, A., M. KOLOS, Y. LIN, C. LIAO, A. S. RIHEL, AND L. ZELLJADT (2021): "Refinitiv Carbon Market Survey 2021," Tech. rep., Refinitiv.

NORDHAUS, W. D. (2017): "Revisiting the social cost of carbon," *Proceedings of the National Academy of Sciences*, 114, 1518–1523.

PESZKO, G., D. VAN DER MENSBRUGGHE, A. GOLUB, J. WARD, D. ZENGHELIS, C. MARIJS, A. SCHOPP, J. A. ROGERS, AND A. MIDGLEY (2020): *Diversification and Cooperation in a Decarbonizing World: Climate Strategies for Fossil Fuel-Dependent Countries*, Washington, DC: World Bank, accepted: 2020-06-30T19:11:44Z.

QUEMIN, S. AND R. TROTIGNON (2021): "Emissions trading with rolling horizons," *Journal of Economic Dynamics and Control*, 125, 104099.

Refinitiv (2019): "Refinitiv Carbon Market Survey 2019," Tech. rep., Refinitiv.

Riahi, K., C. Bertram, D. Huppmann, J. Rogelj, V. Bosetti, A.-M. Cabardos, A. Deppermann, L. Drouet, S. Frank, O. Fricko, S. Fujimori, M. Harmsen, T. Hasegawa, V. Krey, G. Luderer, L. Paroussos, R. Schaeffer, M. Weitzel, B. van der Zwaan, Z. Vrontisi, F. D. Longa, J. Després, F. Fosse, K. Fragkiadakis, M. Gusti, F. Humpenöder, K. Keramidas, P. Kishimoto, E. Kriegler, M. Meinshausen, L. P. Nogueira, K. Oshiro, A. Popp, P. R. R. Rochedo, G. Ünlü, B. van Ruijven, J. Takakura, M. Tavoni, D. van Vuuren, and B. Zakeri (2021): "Cost and attainability of meeting stringent climate targets without overshoot," *Nature Climate Change*, 11, 1063–1069.

Riahi, K., D. P. van Vuuren, E. Kriegler, J. Edmonds, B. C. O'Neill, S. Fujimori, N. Bauer, K. Calvin, R. Dellink, O. Fricko, W. Lutz, A. Popp, J. C. Cuaresma, S. Kc, M. Leimbach, L. Jiang, T. Kram, S. Rao, J. Emmerling, K. Ebi, T. Hasegawa, P. Havlik, F. Humpenöder, L. A. Da Silva, S. Smith, E. Stehfest, V. Bosetti, J. Eom, D. Gernaat, T. Masui, J. Rogelj, J. Strefler, L. Drouet, V. Krey, G. Luderer, M. Harmsen, K. Takahashi, L. Baumstark, J. C. Doelman, M. Kainuma, Z. Klimont, G. Marangoni, H. Lotze-Campen, M. Obersteiner, A. Tabeau, and M. Tavoni (2017): "The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview," *Global Environmental Change*, 42, 153–168.

Salle, I., M. Yildizoğlu, and M.-A. Sénégas (2013): "Inflation targeting in a learning economy: An ABM perspective," *Economic Modelling*, 34, 114–128.

Semieniuk, G., E. Campiglio, J.-F. Mercure, U. Volz, and N. R. Edwards (2021): "Low-carbon transition risks for finance," *Wiley Interdisciplinary Reviews: Climate Change*, 12, e678.

Sendstad, L. H., V. Hagspiel, W. J. Mikkelsen, R. Ravndal, and M. Tveitstøl (2022): "The impact of subsidy retraction on European renewable energy investments," *Energy Policy*, 160, 112675.

Spiro, D. (2014): "Resource prices and planning horizons," *Journal of Economic Dynamics and Control*, 48, 159–175.

Stroebel, J. and J. Wurgler (2021): "What do you think about climate finance?" *Journal of Financial Economics*, 142, 487–498.

Torren-Peraire, D., I. Savin, and J. van den Bergh (2023): "An Agent-Based Model of Cultural Change for a Low-Carbon Transition," .

Train, K. E. (2009): *Discrete Choice Methods with Simulation*, Cambridge: Cambridge University Press, 2 ed.

TRAN, T. T. D. AND A. D. SMITH (2018): "Incorporating performance-based global sensitivity and uncertainty analysis into LCOE calculations for emerging renewable energy technologies," *Applied Energy*, 216, 157–171.

URPELAINEN, J. AND T. VAN DE GRAAF (2018): "United States non-cooperation and the Paris agreement," *Climate Policy*, 18, 839–851.

VAN DEN BREMER, T. S. AND F. VAN DER PLOEG (2021): "The Risk-Adjusted Carbon Price," *American Economic Review*, 111, 2782–2810.

VAN DER PLOEG, F. AND A. REZAI (2020): "The risk of policy tipping and stranded carbon assets," *Journal of Environmental Economics and Management*, 100, 102258.

VOGT-SCHILB, A., G. MEUNIER, AND S. HALLEGATTE (2012): *How Inertia and Limited Potentials Affect the Timing of Sectoral Abatements in Optimal Climate Policy*, Policy Research Working Papers, The World Bank.

VONA, F. (2018): "Job losses and political acceptability of climate policies: why the 'job-killing' argument is so persistent and how to overturn it," *Climate Policy*.

WAISMAN, H., C. GUIVARCH, F. GRAZI, AND J. C. HOURCADE (2012): "The Imaclim-R model: infrastructures, technical inertia and the costs of low carbon futures under imperfect foresight," *Climatic Change*, 114, 101–120.

WORLD BANK (2022): "State and trends of carbon pricing 2022," Tech. rep., The World Bank, Washington, DC.

ZEPPINI, P. (2015): "A discrete choice model of transitions to sustainable technologies," *Journal of Economic Behavior & Organization*, 112, 187–203.

Appendices

Appendix A Microfoundation of logit model

We use a logit model to characterise firms' belief switching and investment choices. This model is based on a discrete choice framework McFadden (1974), which can be microfounded with the random utility framework Train (2009). According to this framework, agents attempt to maximise their utility, which depends on factors common to everyone and explicitly modelled and on idiosyncratic factors treated as random. These idiosyncratic factors can be interpreted as behavioural frictions that impede agents' ability to maximise their utility, leading to heterogeneous choices across agents. In what follows, we show how the random utility model leads to the aggregate belief and investment choices we employ in this paper.

Let us start from the belief switching process, where firm f faces two alternatives $j \in \{b, s\}$, each of which is characterised by a certain random utility Z^*_{fj}. The utility deriving from each alternative is decomposed into a part labelled V_j that is common to all firms, and an idiosyncratic part ϵ^*_{fj} that is treated as random:

$$Z^*_{fj} = V_j + \epsilon^*_{fj} \quad \forall f, j \tag{19}$$

where the common factor $V_j = -\beta^* U_j$ is a function of the fitness measure U_j (see equation (3)) and β^* is the effect of U_j on V_j.

The logit model is obtained by assuming that each ϵ^*_{fj} is independently, identically distributed Gumbel with variance $\sigma^2 \frac{\pi^2}{6}$, where σ is a scale parameter Train (2009). The Gumbel distribution is very similar to a normal, except that it is characterised by slightly fatter tails, thus allowing for slightly more aberrant behaviour than the normal.

By scaling the random utility by $\frac{1}{\sigma}$, the ordering of alternatives is unchanged and the probability of firm f choosing believers' expectation rule is given by the following cumulative density function:

$$P_{fb} = 1 - P_{fs} = \text{Prob}(-\beta U_b + \epsilon_{fb} > -\beta U_s + \epsilon_{fs}), \tag{20}$$

where $\beta = \frac{\beta^*}{\sigma}$ and $\epsilon_{fj} = \frac{\epsilon^*_{fj}}{\sigma}$. The parameter σ thus scales the coefficient β^* to reflect the variance of the idiosyncratic portion of utility. In fact, β indicates the effect of the common variable on the utility relative to the variance of the idiosyncratic factors. A larger variance of the latter factors leads to smaller values of β, even if the common factors have the same effect on utility.

Equation (20) is equivalent to:

$$P_{fb} = \text{Prob}(\epsilon_{fs} - \epsilon_{fb} < \beta U_s - \beta U_b), \tag{21}$$

Since the difference between two Gumbel variables is distributed logistic, then we can express P_{fb}, as well as the share of believers n, as follows:

$$n = P_{fb} = \frac{1}{1 + \exp(-\beta(U_s - U_b))}. \tag{22}$$

Based on expectation rule j, firm f chooses its investment allocation between low- and high-carbon technologies. Both technologies $i \in \{l, h\}$ provide a certain random utility to firm f:

$$Y^*_{fji} = W_{ji} + \epsilon^*_{fji} \tag{23}$$

where $W_{ji} = \gamma^* E_j(\Theta_i)$ is the factor common to all firms with belief type j and is a function of expected production costs of technology i. ϵ^*_{fji} is the factor idiosyncratic to firm f and is assumed iid Gumbel with variance $\sigma^2 \frac{\pi^2}{6}$. Similarly to the belief choice, we scale the random utility by $\frac{1}{\sigma}$ and obtain the probability of firm f with belief j choosing technology l:

$$P_{fjl} = \mathrm{Prob}(\epsilon_{fjh} - \epsilon_{fjl} < \gamma E_j(\Theta_h) - \gamma E_j(\Theta_l)), \tag{24}$$

where $\epsilon_{fji} = \frac{\epsilon^*_{fji}}{\sigma}$ and $\gamma = \frac{\gamma^*}{\sigma}$. Under the assumption that the idiosyncratic factors are iid Gumbel, equation (24) corresponds to:

$$\chi_j = P_{fjl} = \frac{1}{1 + \exp(\gamma(E_j(\Theta_h) - E_j(\Theta_l)))}. \tag{25}$$

Appendix B Proofs and derivation of analytical results

B.1 Proof of Proposition 1

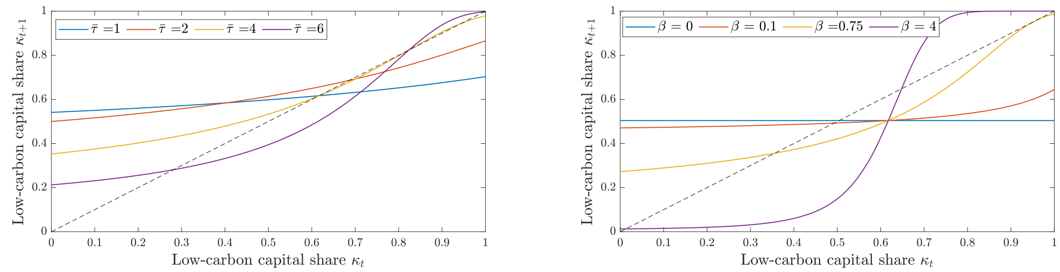
Proof. Let us note that $f(\kappa)$ is continuous in [0, 1] and $f(\kappa) \in [0,1]\ \forall \kappa$. Furthermore, the first derivative $f'(\kappa)$ of the function is given by

$$f'(\kappa) = -\frac{2\,a\,\bar{\tau}^2\,\beta\,\mathrm{e}^{\beta\left[\tau_0 - \bar{\tau}\left(c + \frac{c-1}{a\,\bar{\tau}\,(\kappa-1)-1}\right)\right]}\,\mathrm{e}^{\beta\left[\bar{\tau} - \bar{\tau}\left(c + \frac{c-1}{a\,\bar{\tau}\,(\kappa-1)-1}\right)\right]}\,\tilde{X}_j\,(c-1)}{\left\{\mathrm{e}^{\beta\left[\tau_0 - \bar{\tau}\left(c + \frac{c-1}{a\,\bar{\tau}\,(\kappa-1)-1}\right)\right]}\,\mathrm{e}^{\beta\left[\bar{\tau} - \bar{\tau}\left(c + \frac{c-1}{a\,\bar{\tau}\,(\kappa-1)-1}\right)\right]} + 1\right\}^2 (a\,\bar{\tau} - a\,\bar{\tau}\,\kappa + 1)^2}, \tag{26}$$

where $\tilde{X}_j \equiv \chi_b - \chi_s$.

For $c \neq 1$ and $\beta \neq 0$, $f'(\kappa) > 0\ \forall \kappa \in [0,1]$. It follows that there is at least one stable equilibrium[28]. Moreover, notice that, for finite β and γ, $f(0) = \left[\frac{1}{1+\exp(-\beta(2\tau_t - \tau_0 - \bar{\tau}))}\right](\chi_b - \chi_s) + \chi_s \in (0,1)$ and $f(1) = \left[\frac{1}{1+\exp(-\beta(\bar{\tau} - \tau_0))}\right](\chi_b - \chi_s) + \chi_s \in (0,1)$, which implies that the map starts

[28]As mentioned in section 3, for $c = 1$ and $\beta = 0$, the model features unique steady states.



(a) (b)

Figure A1: Dynamics of the low-carbon share of capital κ for different levels of (a) the tax target $\bar{\tau}$ and (b) belief responsiveness β, under $c = 0.3$.

above the 45 degree line and ends below the 45 degree line. Therefore, generally an overall odd number of steady states exists, excluding those cases where f is tangent to the 45 degree line. □

Figure A1 offers a graphical depiction of the dynamics of the low-carbon share of capital for different values of $\bar{\tau}$ and β, under weak commitment ($c = 0.3$). In both figures, depending on the parameters, we observe the emergence of up to three steady states, identified by the points in which the curves intersect the 45-degrees line. By Proposition 1, the intermediate steady state is unstable while the others are stable. Specifically, figure A1a illustrates that for announced policies that are not too ambitious (e.g. $\bar{\tau} < 4$), the equilibrium κ^* is positive, unique and it increases with the tax target. However, if the policy-maker announces a excessively high tax, two new fixed points emerge. The intermediate steady state defines the boundaries of the basins of attraction of the high- and the low-carbon equilibria. Hence, whenever the intermediate steady state is closer to the low-carbon one, for a policy-maker it is relatively more difficult to obtain a successful low-carbon transition. Figure A1b shows that for low values of belief responsiveness β, the steady state is unique and decreasing in β. As the belief responsiveness crosses a certain threshold (β ¿ 0.75), the steady states become three, where the high- and low-carbon ones are stable and the intermediate one is unstable.

B.2 Proof of Proposition 2

Proof. Under $\beta = \gamma = \infty$, plugging $\kappa_t = 1$ into equation (13), leads to

$$n_{t+1} = \frac{1}{1 + \exp(-\beta(\bar{\tau} - \tau_0))}. \tag{27}$$

Since $\bar{\tau} > \tau_0$ by assumption, it follows that for $\kappa_t = 1$, $n_{t+1} = 1$ and therefore $\kappa_{t+1} = \chi_b$. The believers' low-carbon investment share χ_b can be expressed as[29]

$$\chi_b = \frac{1}{1 + \exp\left(-\gamma \frac{1-D^{R+1}}{1-D}[\theta_h(1+\bar{\tau}) - \theta_l)]\right)}. \tag{28}$$

In order for χ_b to equal 1, under $\gamma = \infty$, the announced tax target $\bar{\tau}$ must make the low-carbon technology more convenient than the high-carbon one ($\bar{\tau} > \frac{\theta_l - \theta_h}{\theta_h}$). If this condition is satisfied, the low-carbon steady state $\kappa_l^* = 1$ exists.

With respect to the high-carbon steady state, let us first note that, under the assumption that $\tau_0 < \frac{\theta_l - \theta_h}{\theta_h}$ and under infinite γ, sceptics do not invest at all in the low-carbon technology ($\chi_s = 0$). Also, from equation (12) it follows that $\kappa = 0$ is a steady state if, for $\kappa_t = 0$, $n_{t+1} = 0$ or $\chi_b = 0$. Concerning the former case ($n_{t+1} = 0$) let us plug $\kappa_t = 0$ into equation (13):

$$n_{t+1} = \frac{1}{1 + \exp\left(-\beta\left[2\bar{\tau}\left(c + \frac{1-c}{1+a\bar{\tau}}\right) - \tau_0 - \bar{\tau}\right]\right)}, \tag{29}$$

which, under infinite β, equals zero if

$$c < \frac{1}{2} - \frac{\bar{\tau} - \tau_0(1 + a\bar{\tau})}{2a\bar{\tau}^2}. \tag{30}$$

Concerning the latter case, $\chi_b = 0$ if $\bar{\tau} < \frac{\theta_l - \theta_h}{\theta_h}$. $\square$

B.3 Proof of Proposition 3

Proof. Concerning the low-carbon steady state, we assume that $\kappa_t = 1 - \tilde{\lambda}_l$, where $\tilde{\lambda}_l = \lambda_l + \varepsilon_l$, with ε_l a sufficiently small positive number, and we impose that $\kappa_{t+1} > 1 - \tilde{\lambda}_l$, meaning that κ is converging to a stable steady state $\kappa_l^* = 1 - \lambda_l$. Hence:

$$\kappa_{t+1} = \frac{\chi_b - \chi_s}{1 + \exp\left(-\beta\left\{2\bar{\tau}\left[c + \left(\frac{1-c}{1+a\tilde{\lambda}_l\bar{\tau}}\right)\right]\right\}\right)} + \chi_s > 1 - \tilde{\lambda}_l, \tag{31}$$

which implies

$$\beta\left\{2\bar{\tau}\left[c + \left(\frac{1-c}{1+a\tilde{\lambda}_l\bar{\tau}}\right)\right]\right\} > -\ln\left(\frac{\chi_b - 1 + \tilde{\lambda}_l}{1 - \tilde{\lambda}_l - \chi_s}\right) \tag{32}$$

[29] After the sum of expected future production costs (equation (6)) in the absence of expected climate policy has been rearranged as $\theta_i \frac{1-D^{R+1}}{1-D}$ and simplified.

and

$$c > \frac{1}{2} - \frac{\bar{\tau} - \tau_0(1 + a\tilde{\lambda}_l\bar{\tau})}{2a\tilde{\lambda}_l\bar{\tau}^2} - \ln\left(\frac{\chi_b - 1 + \tilde{\lambda}_l}{1 - \tilde{\lambda}_l - \chi_s}\right)(2\bar{\tau}\beta)^{-1}\left(1 + \frac{1}{a\tilde{\lambda}_l\bar{\tau}}\right). \tag{33}$$

In order for equation (33) to be well defined, we impose

$$\begin{aligned} \chi_b &> 1 - \tilde{\lambda}_l, \\ \chi_s &< 1 - \tilde{\lambda}_l. \end{aligned} \tag{34}$$

The former condition is satisfied if

$$\bar{\tau} > \frac{\theta_l - \theta_h}{\theta_h} - \frac{\ln\left(\frac{\tilde{\lambda}_l}{1-\tilde{\lambda}_l}\right)}{\frac{1-D^{R+1}}{1-D}\gamma\theta_h}, \tag{35}$$

where $D \equiv \frac{1}{1+\rho}$. The second condition is satisfied if

$$\tau_0 < \frac{\theta_l - \theta_h}{\theta_h} - \frac{\ln\left(\frac{\tilde{\lambda}_l}{1-\tilde{\lambda}_l}\right)}{\frac{1-D^{R+1}}{1-D}\gamma\theta_h}. \tag{36}$$

Since we assume that $\tau_0 < \frac{\theta_l - \theta_h}{\theta_h}$, if $\tilde{\lambda}_l < 0.5$, then condition 36 is always verified. If, instead, $\tilde{\lambda}_l > 0.5$, this is an additional constraint.

Concerning the high-carbon steady state, we assume that $\kappa_t = \chi_s + \tilde{\lambda}_h$, where $\tilde{\lambda}_h = \lambda_h + \varepsilon_h$, with ε_h a sufficiently small positive number, and we impose that $\kappa_{t+1} < \chi_s + \tilde{\lambda}_h$, meaning that κ is converging to a stable steady state $\kappa_h^* = \chi_s + \lambda_h$. Hence:

$$\kappa_{t+1} = \frac{\chi_b - \chi_s}{1 + \exp\left(-\beta\left[2\bar{\tau}\left(c + \frac{1-c}{1+a(1-\chi_s-\tilde{\lambda}_h)\bar{\tau}}\right)\right]\right)} + \chi_s < \chi_s + \tilde{\lambda}_h \tag{37}$$

which implies

$$c < \frac{1}{2} + \frac{\bar{\tau} - \tau_0\{1 + a[1 - (\chi_s + \tilde{\lambda}_h)]\bar{\tau})\}}{2a[1 - (\chi_s + \tilde{\lambda}_h)]\bar{\tau}^2} - \frac{\ln\left(\frac{\chi_b - \chi_s - \tilde{\lambda}_h}{\tilde{\lambda}_h}\right)}{2\bar{\tau}\beta}\left\{1 + \frac{1}{a[1 - (\chi_s + \tilde{\lambda}_h)]\bar{\tau}}\right\}. \tag{38}$$

In order for equation (38) to be well defined, we impose $\tilde{\lambda}_h < \chi_b - \chi_s$.

□

B.4 Proof of Proposition 4

Proof. The second derivative of $f(\kappa)$ is:

$$f''(\kappa) = -\frac{\tilde{x}\,\mathrm{e}^{\tilde{\beta}}\,\Big[(a\,\bar{\tau} - \bar{\tau}\,\beta + \bar{\tau}\,\beta\,c - a\,\bar{\tau}\,\kappa_t + 1) + \mathrm{e}^{\tilde{\beta}}\,(a\,\bar{\tau} + \bar{\tau}\,\beta - \bar{\tau}\,\beta\,c - a\,\bar{\tau}\,\kappa_t + 1)\Big]}{\left(\mathrm{e}^{\tilde{\beta}} + 1\right)^3 (a\,\bar{\tau} - a\,\bar{\tau}\,\kappa_t + 1)^4}, \tag{39}$$

where $\tilde{x} \equiv (\chi_b - \chi_s)\,(c-1)\,4\,a^2\,\bar{\tau}^3\,\beta < 0$ and $\tilde{\beta} \equiv \beta\,(\tau_0 - 2\tau_t + \bar{\tau})$. Although we cannot find analytically the inflection points of κ where $f''(\kappa) = 0$, we observe that, for $\beta \neq 0$ and $c \neq 1$, if $c > 1 - \frac{1}{\bar{\tau}\beta}$, then $(a\,\bar{\tau} - \bar{\tau}\,\beta + \bar{\tau}\,\beta\,c - a\,\bar{\tau}\,\kappa_t + 1) > 0$ and $f''(\kappa) > 0$ for all $\kappa \in [0,1]$ and the fixed point is unique. If $c < 1 - \frac{1}{\bar{\tau}\beta}$, one or more fixed points may exist. □

Appendix C Sensitivity analysis

Table A1: Sensitivity analysis

	Full commitment ($c = 1$)				Poor commitment ($c = 0.3$)			
Parameter	κ in 2050	κ in 2100	n in 2050	n in 2100	κ in 2050	κ in 2100	n in 2050	n in 2100
Production								
Output growth rate g_Y								
0.003	0.5668	0.9842	0.6923	1.0000	0.5212	0.1561	0.5953	0.0004
0.005	0.5774	0.9882	0.6923	1.0000	0.5302	0.1392	0.5967	0.0004
0.007	0.5864	0.9912	0.6923	1.0000	0.5379	0.1239	0.5978	0.0004
0.009	0.5942	0.9933	0.6923	1.0000	0.5443	0.1101	0.5988	0.0004
0.011	0.6008	0.9949	0.6923	1.0000	0.5498	0.0979	0.5996	0.0004
Depreciation rate δ								
0.006	0.4732	0.9172	0.6923	1.0000	0.4397	0.2624	0.5840	0.0005
0.01	0.5206	0.9588	0.6923	1.0000	0.4812	0.2176	0.5896	0.0005
0.014	0.5538	0.9784	0.6923	1.0000	0.5100	0.1754	0.5937	0.0004
0.018	0.5774	0.9882	0.6923	1.0000	0.5302	0.1392	0.5967	0.0004
0.022	0.5946	0.9934	0.6923	1.0000	0.5447	0.1094	0.5988	0.0004
Initial low-carbon capital share κ_0								
0.1	0.5709	0.9882	0.6923	1.0000	0.5228	0.1385	0.5956	0.0004
0.2	0.5774	0.9882	0.6923	1.0000	0.5302	0.1392	0.5967	0.0004
0.3	0.5839	0.9883	0.6923	1.0000	0.5377	0.1398	0.5978	0.0004
0.4	0.5904	0.9884	0.6923	1.0000	0.5451	0.1404	0.5989	0.0004
0.5	0.5969	0.9884	0.6923	1.0000	0.5525	0.1410	0.6001	0.0004
Low- to high-carbon production costs $\frac{\theta_l}{\theta_h}$								
1	0.9470	0.9995	0.6923	1.0000	0.9470	0.9995	0.6780	1.0000
1.1	0.9212	0.9987	0.6923	1.0000	0.9203	0.9986	0.6713	1.0000
1.2	0.6066	0.9891	0.6923	1.0000	0.5651	0.2158	0.6019	0.0005
1.3	0.5812	0.9883	0.6923	1.0000	0.5345	0.1399	0.5973	0.0004
1.4	0.5719	0.9882	0.6923	1.0000	0.5241	0.1386	0.5957	0.0004
1.5	0.5552	0.9880	0.6923	1.0000	0.5056	0.1372	0.5930	0.0004
Investment responsiveness γ								

Table A1: Sensitivity analysis

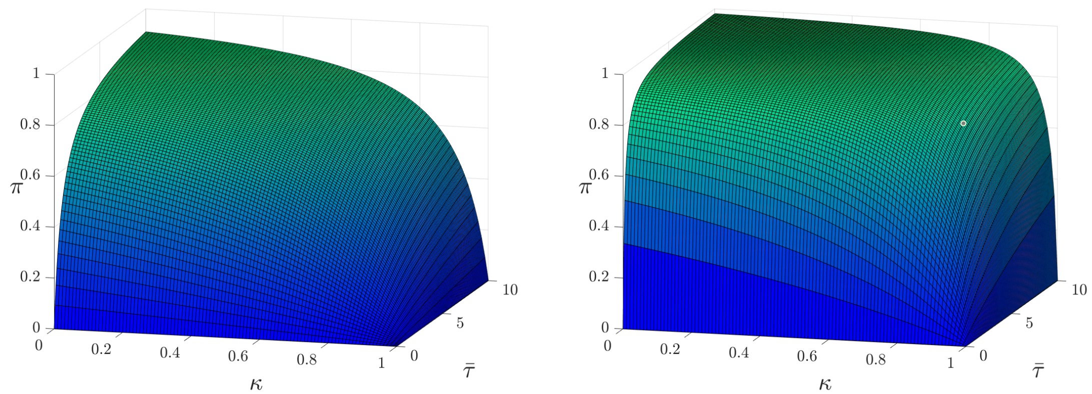
	Full commitment ($c = 1$)				Poor commitment ($c = 0.3$)			
Parameter	κ in 2050	κ in 2100	n in 2050	n in 2100	κ in 2050	κ in 2100	n in 2050	n in 2100
0	0.4805	0.4998	0.6923	1.0000	0.4805	0.4998	0.5897	0.0011
1	0.5774	0.9882	0.6923	1.0000	0.5302	0.1392	0.5967	0.0004
2	0.5781	0.9883	0.6923	1.0000	0.5311	0.1392	0.5968	0.0004
3	0.5783	0.9883	0.6923	1.0000	0.5312	0.1392	0.5968	0.0004
4	0.5783	0.9883	0.6923	1.0000	0.5312	0.1392	0.5968	0.0004
Planning horizon R								
80	0.5634	0.9881	0.6923	1.0000	0.5147	0.1380	0.5943	0.0004
100	0.5716	0.9882	0.6923	1.0000	0.5237	0.1386	0.5957	0.0004
120	0.5774	0.9882	0.6923	1.0000	0.5302	0.1392	0.5967	0.0004
140	0.5805	0.9883	0.6923	1.0000	0.5337	0.1394	0.5972	0.0004
160	0.5813	0.9883	0.6923	1.0000	0.5346	0.1395	0.5973	0.0004
Discount rate ρ								
0.01	0.5811	0.9883	0.6923	1.0000	0.5344	0.1395	0.5973	0.0004
0.02	0.5747	0.9882	0.6923	1.0000	0.5273	0.1389	0.5962	0.0004
0.03	0.5648	0.9881	0.6923	1.0000	0.5162	0.1389	0.5946	0.0004
0.04	0.5558	0.9881	0.6923	1.0000	0.5064	0.1418	0.5931	0.0004
Expectations								
Belief responsiveness β								
0	0.4755	0.4998	0.5000	0.5000	0.4755	0.4998	0.5000	0.5000
1	0.5774	0.9882	0.6923	1.0000	0.5302	0.1392	0.5967	0.0004
2	0.6644	0.9951	0.8351	1.0000	0.5877	0.9865	0.7016	1.0000
3	0.7306	0.9969	0.9193	1.0000	0.6442	0.9942	0.8017	1.0000
4	0.7780	0.9977	0.9625	1.0000	0.6953	0.9962	0.8821	1.0000
Memory parameter η								
0	0.5795	0.9885	0.6956	1.0000	0.5314	0.1351	0.5980	0.0004
0.2	0.5790	0.9884	0.6947	1.0000	0.5311	0.1361	0.5977	0.0004
0.4	0.5781	0.9883	0.6934	1.0000	0.5306	0.1378	0.5971	0.0004

Table A1: Sensitivity analysis

	Full commitment ($c = 1$)				Poor commitment ($c = 0.3$)			
Parameter	κ in 2050	κ in 2100	n in 2050	n in 2100	κ in 2050	κ in 2100	n in 2050	n in 2100
0.6	0.5763	0.9881	0.6908	1.0000	0.5297	0.1412	0.5960	0.0004
0.8	0.5710	0.9875	0.6832	1.0000	0.5266	0.1511	0.5926	0.0006
1	0.2900	0.2994	0.2994	0.2994	0.2900	0.2994	0.2994	0.2994
Policy								
Initial tax rate τ_0								
0.05	0.3606	0.9454	0.5672	0.9995	0.3523	0.5940	0.5481	0.3813
0.15	0.5774	0.9882	0.6923	1.0000	0.5302	0.1392	0.5967	0.0004
0.25	0.6427	0.9939	0.7944	1.0000	0.5393	0.0605	0.5982	0.0000
0.35	0.7874	0.9973	0.8690	1.0000	0.7070	0.4225	0.6554	0.0000
Tax target growth rate $\bar{g}_\tau$								
0.005	0.0418	0.6025	0.5297	0.6408	0.0408	0.5575	0.5104	0.5817
0.01	0.4723	0.8864	0.5819	0.9666	0.4519	0.6521	0.5473	0.6586
0.015	0.5628	0.9847	0.6696	1.0000	0.5218	0.2049	0.5884	0.0053
0.02	0.6343	0.9946	0.7977	1.0000	0.5530	0.0424	0.6242	0.0000
0.025	0.7124	0.9969	0.9274	1.0000	0.5682	0.0161	0.6253	0.0000
0.03	0.7820	0.9979	0.9902	1.0000	0.5571	0.0084	0.5221	0
Transition risk index parameter a								
0.5	0.5774	0.9882	0.6923	1.0000	0.5513	0.9736	0.6391	1.0000
1	0.5774	0.9882	0.6923	1.0000	0.5302	0.1392	0.5967	0.0004
1.5	0.5774	0.9882	0.6923	1.0000	0.5134	0.0859	0.5635	0.0002
2	0.5774	0.9882	0.6923	1.0000	0.4998	0.0688	0.5376	0.0002
2.5	0.5774	0.9882	0.6923	1.0000	0.4886	0.0603	0.5172	0.0002
3	0.5774	0.9882	0.6923	1.0000	0.4794	0.0552	0.5008	0.0001

Appendix D Transition risk index function: graphical representation and robustness

First, let us provide a graphical representation of the transition risk index function (see figure A2).



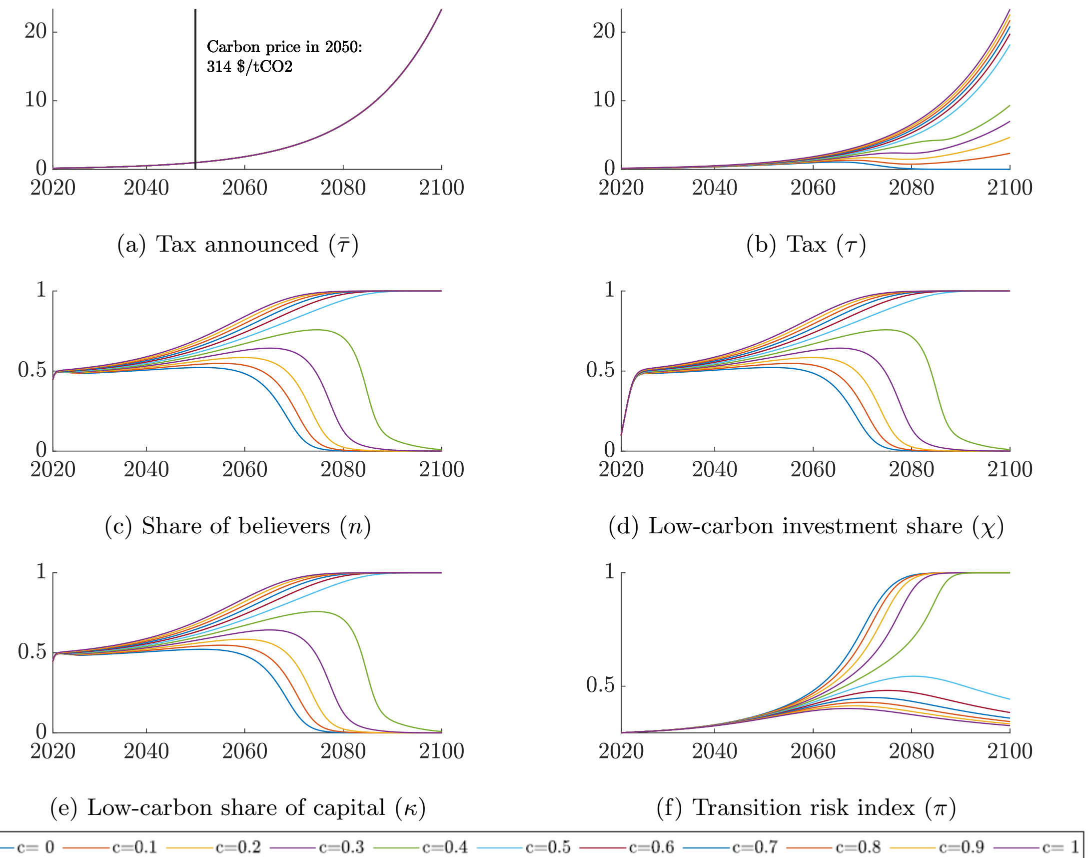
(a) $a = 1$ (b) $a = 5$

Figure A2: Transition risk index π as a function of κ and $\bar{\tau}$, for two distinct levels of a.

Second, we present the transition path of the model under the assumption of a logistic transition risk index function, defined as

$$\pi_t = \frac{1}{1 + \exp\left(-a[(1-\kappa_t)\bar{\tau}_t - \pi_0]\right)}, \tag{40}$$

where π_0 is the inflection point of the function and is calibrated here to 1. As shown in figure A3, the dynamics of the model are qualitatively similar to the baseline version of the model.



(a) Tax announced ($\bar{\tau}$)

(b) Tax (τ)

(c) Share of believers (n)

(d) Low-carbon investment share (χ)

(e) Low-carbon share of capital (κ)

(f) Transition risk index (π)

Figure A3: Evolution over time of selected variables under various levels of commitment (c), assuming a logistic transition risk index function.